

STIC-Biotech/ChemLib

From: Chan, Christina  
Sent: Tuesday, December 04, 2001 4:01 PM  
To: Davis, Minh-Tam; STIC-Biotech/ChemLib  
Subject: RE: Rush search request for 09/555342

Please rush. Thanks Chris

-----Original Message-----

From: Davis, Minh-Tam  
Sent: Tuesday, December 04, 2001 3:21 PM  
To: Chan, Christina  
Subject: Rush search request for 09/555342

Please search in commercial database and in issued patent files:

- 1) SEQ ID NOS: 1 and 2
- 2) Amino acids 1-374 of SEQ ID NO:2.
- 3) Amino acids 544-737 of SEQ ID NO:2
- 4) Amino acids 764 to 854 of SEQ ID NO:2.
- 5) Nucleotides 49-3183 of SEQ ID NO:1.
- 6) Oligomer search for at least more than 90 amino acids of SEQ ID NO:2

Thank you,  
MINH TAM DAVIS  
ART UNIT 1642, ROOM 8A01, MB 8E12  
305-2008

*priority date:*

*Nov 97*

*CRFF*

if Contact:  
Sheppard

Searcher: \_\_\_\_\_  
Phone: tel: 308-4499  
Location: \_\_\_\_\_  
Date Picked Up: \_\_\_\_\_  
Date Completed: 12/7/01  
Searcher Prep/Review: \_\_\_\_\_  
Clerical: \_\_\_\_\_  
Online time: \_\_\_\_\_

TYPE OF SEARCH:

NA Sequences: \_\_\_\_\_  
AA Sequences: \_\_\_\_\_  
Structures: \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Full text: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

VENDOR/COST(where applic.)

STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
Questel/Orbit: \_\_\_\_\_  
DRLink: \_\_\_\_\_  
Lexis/Nexis: \_\_\_\_\_  
Sequence Sys.: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other (specify): \_\_\_\_\_

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw.model

Run on: December 6, 2001, 08:46:08 ; Search time 72.4 Seconds  
(without alignments)  
1069.152 Million cell updates/sec

Title: US-09-555-342A-2

Perfect score: 5463

Sequence: 1 MGEIQRPTPGSLGAPENS.....SATSSASRPHVLSHKESLVY 1045

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq.ll101.\*

1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	5463	100.0	1045	20	Human chondrocyte-
2	5459	99.9	1045	21	Human cytoskeleton
3	1737	31.8	395	21	Human pancreatic c
4	720	13.2	1003	21	Human cytoskeleton
5	626	11.5	687	22	Human cytoskeleton
6	541	9.9	913	18	Human protein sequ
7	541	9.9	913	20	Protein tyrosine p
8	520.5	9.5	766	21	Human ptpH1 protei
9	419	7.7	369	22	Rat actin-binding
10	409	7.5	432	22	Human protein tyro
11	396	7.2	1174	19	Human protein sequ

12	396	7.2	1174	20	AAV34158	Human protein tyro
13	367.5	6.7	635	21	AAV33356	Human colon cancer
14	356.5	6.5	586	20	AAV27443	Amino acid sequenc
15	342.5	6.3	305	22	AAV25892	Human protein sequ
16	335	6.1	590	20	AAV94458	Human neurofibroma
17	335	6.1	591	18	AAV09648	Human merlin prote
18	335	6.1	595	15	AAV60398	Merlin protein enc
19	335	6.1	596	20	AAV94459	Human neurofibroma
20	334	6.1	584	20	AAV94457	Mouse neurofibroma
21	334	6.1	591	20	AAV94456	Mouse neurofibroma
22	334	6.1	596	20	AAV94455	Mouse neurofibroma
23	333	6.1	495	22	AAV41770	Human polypeptide
24	332	6.1	584	18	AAV09647	Mouse merlin prote
25	330	6.0	591	18	AAV09646	Mouse merlin prote
26	330	6.0	596	18	AAV09645	Mouse merlin prote
27	320.5	5.9	445	22	AAV39984	Human polypeptide
28	308	5.6	2485	21	AAV19343	Amino acid sequenc
29	306	5.6	279	22	AAV25733	Human protein sequ
30	300.5	5.5	2466	16	AAV71498	Human protein tyro
31	300.5	5.5	2466	19	AAV75999	Intracellular prot
32	300.5	5.5	2466	21	AAV90272	Human ptpH1 phosph
33	287.5	5.3	450	22	AAV92607	Human protein sequ
34	266	4.9	1105	22	AAV25567	Human protein sequ
35	264.5	4.8	2861	18	AAV27227	Human TRIO phospho
36	256.5	4.7	619	22	AAV97025	Human colon carcin
37	245	4.5	142	22	AAV74414	Human colon cancer
38	244.5	4.5	1715	21	AAV57449	Mouse Ees1L protei
39	242	4.4	1520	20	AAV41010	Amino acid sequenc
40	240.5	4.4	1527	22	AAU01184	Rat glutamate tran
41	239	4.4	1683	21	AAV71160	Rat phosphodiester
42	235.5	4.3	523	22	AAV39338	Human polypeptide
43	234.5	4.3	647	22	AAV41124	Human polypeptide
44	227.5	4.2	580	20	AAV81349	Human guanine nucl
45	223	4.1	1658	21	AAV57450	Mouse Ees2L protei

#### ALIGNMENTS

RESULT 1  
AAV07482  
ID AAV07482 standard; Protein; 1045 AA.  
XX AC AAV07482;  
XX DF 17-AUG-1999 (first entry)  
XX DE Human chondrocyte-derived protein CDEP.  
XX KW Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;  
KW Dbl homology domain; plectstrin homology domain; rheumatoid 'arthritis;  
KW drug.  
XX OS Homo sapiens.  
XX PN WO9928458-A1.  
XX PD 10-JUN-1999.  
XX PF 27-NOV-1998; 98WO-JP05348.  
XX PR 27-NOV-1997; 97JP-0342060.  
XX (CHUS ) CHUGAI SEIYAKU KK.  
XX PI Kato Y, Kawamoto T, Koyano Y;  
XX WPI; 1999-371117/31.  
XX N-PSDB; AAV79183.  
XX PT Protein CDEP expressed in differentiated chondrocytes, and gene  
XX encoding it

PS Claim 2: Fig 1; 59pp; Japanese.

XX This sequence represents a protein (CDEP) expressed in differentiated human foetal chondrocytes, which contains an erin-like domain, a Db1 CC homology (DH) domain and a pleckstrin homology (PH) domain. The encoding CC nucleic acid or protein can be used in the investigation and treatment of CC cancers and arthritic diseases (including chronic rheumatoid arthritis), CC or for screening of candidate anticancer drugs.

XX Sequence 1045 AA;

Query Match 100.0%; Score 5463; DB 20; Length 1045;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1045; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGEIQRPTPGSLGAPENSGISTLERGQKPPPTPSGKLVSIKIQMLDDTQAFVQRA 60  
Db 1 mgeieqrptpgslgapensgistlergqkppptpsgklvsiqimldtqafevpgra 60  
QY 61 PGKVLDAVCNHLNVEGDYFGLERPDHKKITVWLDLLKPIVKQIRRPKHVVVVFVVKFF 120  
Db 61 pgkvlldavcnhl nvegd yfglerpdhkkitvwldllkpi vkqirrpkhvvvkvvkvff 120  
QY 121 PDHTQLQELTRYLPALQVQDLAQRGRLTNDTSAAILLISHIVQSEIGDFDEALDREHL 180  
Db 121 pdhtqlqeeltrylfalqvqdlagrltndtsaallishivqseigdfdealdrehl 180  
QY 181 AKNKYIPQDALEDKIVEFHNNHIGTTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKI 240  
Db 181 aknkyipqdaledkivefhnnhigtptaesdfolleiarrlemygir lhpakdregtki 240  
QY 241 NLAVANTGILVFGFTKINAFNAWVKLSFKRKRFLIKLRPDANSAYODTLEFLMASRD 300  
Db 241 nlavantgilvfgftkinafnawvklsfkrkrflilkrpdansayodtleflmasrd 300  
QY 301 FCKSEWKICVEHHAFFRLEPEKPKPVLFSGSGSFRFSGRTQKQVLDYVREGGKKVQ 360  
Db 301 fcksewkicvehhafrlepekpkpvl fsgsgsfrfsgrtqkvldy vreggk kvq 360  
QY 361 FERKSHKHSIRSLASQPELNVSEVLEQSQSTSLTFGGAESPQSGQSCRKEPKVSAG 420  
Db 361 ferkshkhsirslasqpeelnvsevl eqsqstsltf ggaespgsgqscrkepkvsag 420  
QY 421 EPQSHSPAPRSPAGNKQADGAASAPTEEEVEVWDRTOQSKPQPOPSTGSLTGSPHL 480  
Db 421 epqshspaprrspagnkqadgaasapteevevwdrtqqskpqp ppsstgsltgsphl 480  
QY 481 SELSVNSQGGVAPANYTLSPNLSPTKQASPLISPLLNDOACPRDDEDEGRKRKPTDK 540  
Db 481 selsvnsqggvapanytlspnlspdkqasplispllndqacprtdedegrkrkptdk 540  
QY 541 AYFIAKEVSTERTYIKDLVITSWFQSVSKEDAMPEALKSLIPNFPPLHKFHTNFKL 600  
Db 541 ayfiak evstertyikdlvitswfskvedampealkslipnfpplhk fhtnflk 600  
QY 601 ETEORLALWEGRSNAQIRDYQRIGDVMKLNIOGCMKHLAHLWKHSALALENGIKSSRR 660  
Db 601 eteoralwegrснаqirdyqrigdvm lknio gcmkhlahlwkhsalealengikssrr 660  
QY 661 LENFCRDFELQVCYLPLNTFLLRPLRLMHYKQVLERLCKHHPPSHADFRDCRAALAEI 720  
Db 661 lenfcrdfelqvcylplntfllrplrlmhykvlerlckhhppshadfrdcraalaei 720  
QY 721 TENVAQLHGTMIKMFQKLKDLIGDNLVPGREFIRLGSLSKLSGKGLQORMFF 780  
Db 721 tenvaqlhgtmikmfqklkdlig dnlvpgrefir lgslskls gkglqormff 780  
QY 781 LFNVDVLLYTSRGLTAGNQFKHQLPLYGMTTEESDEWGVPHCLTLRGOROSIIVAAS 840  
Db 781 lfnvdvlllytsrgltagnqfk hqlplygmtteesde wgvphcltlrgorosiivaas 840  
QY 841 RSEMEKWVEDIOWAIDLAEKSSPAPFLASSPPDNKSPDEATAADQESDEDDLSASRTSL 900

Db 841 remekwvedigwaidlaeksspapflassppdnkspdeataadqesddlsasrtsl 900  
QY 901 ERQAPHRGNTMWHVCHRTSYSMVDFSTAVENQLSGNLLRKFKNSNGWOKLWVFTNFC 960  
Db 901 erqaphrgntmwhcvhrtysvmvdfstavenqlsgnllrkfknsngwqklwvftnfc 960  
QY 961 LFFYKSHQDNHPLASLPLGYSLTIPSESENQKQVFKLHFKSHVYFRASEYTFERW 1020  
Db 961 lffy kshqdnhplaslpllgysltipseseni qkvfklhfkshvyv fraesityferw 1020  
QY 1021 MEVIRSATSSASRPHVLSHESLVY 1045  
Db 1021 mevirsatssasrphvlskheslv y 1045  
RESULT 2  
AA91947  
ID AA91947 standard; Protein; 1045 AA.  
XX  
AC AA91947;  
XX  
DT 19-JUL-2000 (first entry)  
XX  
DE Human cytoskeleton associated protein 2 (CYSKP-2).  
KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;  
KW autoimmunity; inflammatory; vesicle trafficking; neurological;  
KW cardiovascular; cell motility; reproductive; muscle disorder.  
XX  
OS Homo sapiens.  
FH Key Location/Qualifiers  
FT Modified-site 23 /note= "potential phosphorylation site"  
FT Modified-site 36 /note= "potential phosphorylation site"  
FT Modified-site 41 /note= "potential phosphorylation site"  
FT Domain 47..85 /note= "signature sequence"  
FT Modified-site 92 /note= "potential phosphorylation site"  
FT Domain 94..123 /note= "signature sequence"  
FT Domain 144..190 /note= "signature sequence"  
FT Modified-site 150 /note= "potential phosphorylation site"  
FT Modified-site 152 /note= "potential N-glycosylation site"  
FT Domain 196..249 /note= "signature sequence"  
FT Modified-site 207 /note= "potential phosphorylation site"  
FT Domain 261..279 /note= "signature sequence"  
FT Modified-site 270 /note= "potential phosphorylation site"  
FT Modified-site 336 /note= "potential phosphorylation site"  
FT Modified-site 340 /note= "potential phosphorylation site"  
FT Modified-site 343 /note= "potential phosphorylation site"  
FT Modified-site 366 /note= "potential phosphorylation site"  
FT Modified-site 370 /note= "potential phosphorylation site"  
FT Modified-site 396 /note= "potential phosphorylation site"  
FT Modified-site 408 /note= "potential phosphorylation site"  
FT

FT Modified-site 418 /note= "potential phosphorylation site"  
FT Modified-site 448 /note= "potential phosphorylation site"  
FT Modified-site 495 /note= "potential N-glycosylation site"  
FT Modified-site 525 /note= "potential phosphorylation site"  
FT Modified-site 538 /note= "potential phosphorylation site"  
FT Modified-site 542 /note= "potential phosphorylation site"  
FT Modified-site 549 /note= "potential phosphorylation site"  
FT Modified-site 551 /note= "potential phosphorylation site"  
FT Modified-site 571 /note= "potential phosphorylation site"  
FT Modified-site 657 /note= "potential phosphorylation site"  
FT Modified-site 658 /note= "potential phosphorylation site"  
FT Modified-site 706 /note= "potential phosphorylation site"  
FT Modified-site 770 /note= "potential phosphorylation site"  
FT Modified-site 789 /note= "potential phosphorylation site"  
FT Modified-site 811 /note= "potential phosphorylation site"  
FT Modified-site 815 /note= "potential phosphorylation site"  
FT Modified-site 826 /note= "potential phosphorylation site"  
FT Modified-site 839 /note= "potential phosphorylation site"  
FT Domain 770..773 /note= "signature sequence"  
FT Modified-site 840 /note= "potential phosphorylation site"  
FT Modified-site 842 /note= "potential phosphorylation site"  
FT Modified-site 872 /note= "potential phosphorylation site"  
FT Modified-site 878 /note= "potential phosphorylation site"  
FT Modified-site 883 /note= "potential phosphorylation site"  
FT Modified-site 889 /note= "potential phosphorylation site"  
FT Modified-site 898 /note= "potential phosphorylation site"  
FT Modified-site 919 /note= "potential N-glycosylation site"  
FT Modified-site 923 /note= "potential phosphorylation site"  
FT Modified-site 966 /note= "potential phosphorylation site"  
XX WO200017355-A2.  
XX 30-MAR-2000.  
XX 17-SEP-1999; 99WO-US21565.  
XX 18-SEP-1998; 98US-0172226.  
XX 27-APR-1999; 99US-0131321.  
XX (INCY-) INCYTE PHARM INC.  
XX Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;  
PI Guegler KJ, Patterson C, Azimzai Y, Baughn MR;  
XX

DR WPI: 2000-283582/24.  
DR N-PSDB: AAA08582.  
XX Human cytoskeleton associated proteins, used to treat cell  
PT proliferative, autoimmune/inflammatory, vesicle trafficking,  
PT neurological, cell motility, reproductive and muscle disorders  
XX Claim 1; Page 82-84; 113pp; English.  
PS AAY91946-61 show human cytoskeleton associated proteins 1 to 16 (CYSKP-1  
XX to CYSKP-16) respectively. The sequences can be used to treat and  
CC diagnose cancer and cell proliferative, autoimmune/inflammatory, vesicle  
CC trafficking, neurological, cardiovascular, cell motility, reproductive  
CC and muscle disorders. Pharmaceutical compositions containing CYSKP-2 to  
CC CYSKP-16 can be used to treat or prevent disorders associated with  
CC decreased expression or activity of CYSKP (claimed), for example,  
CC atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,  
CC cancers, autoimmune/antiinflammatory disorders such as allergies, anemia,  
CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,  
CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma  
CC and trauma. CYSKP antagonists can be used to treat or prevent a  
CC disorder associated with increased expression or activity of CYSKP  
CC (claimed).  
XX Sequence 1045 AA;  
SQ  
Query Match 99.9%; Score 5459; DB 21; Length 1045;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1044; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 MGEIQRPPTGSRGAPENSIGSTLERGQKPPPTSGKLVSITKIQMLDQTQAFVPPQRA 60  
DB 1 mgeieqrptpgsrlgapensgistlergqkppptsgklvsikigmlddtgeafvppqra 60  
QY 61 PGKVLDDAVCNHLNVEGDYFGLFEPDHHKITVMDLLKPIVKQIRRRPKHVVKVVF 120  
DB 61 pgkvlldavcnhlnvegdylfglepdkhkitvmdllkpiwkqirrvpkhvvkvvkvff 120  
QY 121 PPDHTQLQBELTRYLFALQVKODLAQGRUTCNDTSAALLISHIVQSEIGDFDDEALDREHL 180  
DB 121 ppdhtqlqeeltrylfalqvkdlaqgrlucndtsaallishivqseigdfdealdrehl 180  
QY 181 ANKYIPODDALEDKIVEFHNNHIGQTPAESDFOLLEIARRLMEYGIURLHPAKDREGTKI 240  
DB 181 ankypqgdaldedkivefhnnhigqtpaesdfqlleiarriemylgrrlhpakdregtki 240  
QY 241 NLAVANTGILVFGQETKINAFNWKVKRLSKFRKFLKLRPDANSAYODTLEFLMASRD 300  
DB 241 nlavantgilvfggtklnafnwkvrlskfrkflklrpdansayodtleflmasrd 300  
QY 301 FCKSFWKICVEHAFRFLFPEPKPKPVLFSGSGSFRSGRTQKQVLDYVKEGGHKVQ 360  
DB 301 fcksfwkicvehahfrflfpeepkpkpvlfsrgsgsfrsgrtqkqvldyvkeggkhkvq 360  
QY 361 FERKHSKIHISIRSLASQPTLNSEVLEQSQSTSTTFGEAESPGQSCRRKEPKVKSAG 420  
DB 361 ferkhskihsirslasqptelnsevlseqsststtfgeaespgqscrrkepkvksag 420  
QY 421 EPGSHPSAPRSPAGNKQADGAASAPTEEEVVKDRTQOSKPPQPPQSTGSLTGSPLH 480  
DB 421 epgshpsaprrspagnkqadgaasapteeeevvkdrtooskppppqstgsltgsphl 480  
QY 481 SELSVNSGGVAPANVTLSPLNSPDTKQASPLISPLLNDAQPRTDDEGRKKRPPTDK 540  
DB 481 selsvnsqgvapantvltspnlspdtkqasplispllndaqprtddedegrrkrfptdk 540  
QY 541 AYFIAKEVSTTERTYIKDLEVITSWFQSTVSKEDAMPEALKSLIFPNFPLHKFHTNFK 600  
DB 541 ayfiakevsttertyikdleviswfsqstvsksedampealkslifpnfplhkhtnflk 600  
QY 601 ETEQRIALMEGRSNAQIRDYQIRIGDVMLKNIQGMKHLAAHLWKHSALEALENGKSSRR 660  
DB 601 eteqrialmegrsnaqirdyqirigdvmlkniqgmkhlaahlwkhsalealengkssrr 660

Db 601 eiegrlalwegrnaqirdyqrigdvmlknigmkhlaahlwkhsealealengikssrr 660  
QY 661 LENFCRDFELQKCYLPLNTFLLRLHLMHYKQVLERLCKHPPSHADFRDCRAALAEI 720  
Db 661 lenfcdrfelqkvcyplnflrlphrhmhykqvierickhppshadfrdcraael 720  
QY 721 TEMVAQLHGTMIKMFNQKHLKDLIGDNLVVPGRFIRIGSLSKSGKGLQQRMF 780  
Db 721 temvaqlhgtmikenmfqkhlkdlidnlvvpgrfirlgslsksgkqlqrmff 780  
QY 781 LFNDVLLYTSRGLTASNQFVHGOLPLVGMTESEDEWGVPHCLTLRGQRQIIVAAS 840  
Db 781 lfndvlltysrgltasngfkhgplvgmtieesedewgvphcltlrgqrqsiivaas 840  
QY 841 RSEMEKVEDIQAIDLAEKSSSPAPFLASSPPDNKSPDEATAADQESDDLSASRTSL 900  
Db 841 rsemekevdiqaidlaekssspapflassppdnkspdeataadqesddlsasrtsl 900  
QY 901 ERQAPHRGNTVMVCHRNNTSVSMVDFSTAVENQLSGNLLRKFNKNGMOKLWVFTNFC 960  
Db 901 erqaphrgntvmvchrnntsvsmvdfstavenqlsgnllrkfnkngmqlwvvtncf 960  
QY 961 LFYFKSHQDNHPLASPLLCYSITIPSESENIQKDYVFKLHFKSHVYVYFRAESEYTFERW 1020  
Db 961 lfykshqdnhplaslpllgysitipseesenikdyvfkhlkshvyvfyraeseytferw 1020  
QY 1021 MEVIRSATSSASRPHVLSHKSLEY 1045  
Db 1021 mevirsatssasrphvlsklesivy 1045

RESULT 3  
AAB54227  
ID AAB54227 standard; Protein; 395 AA.  
XX AAB54227;  
AC AAB54227;  
XX AAB54227;  
DT 09-MAR-2001 (first entry)  
DE Human pancreatic cancer antigen protein sequence SEQ ID NO:679.  
XX Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
KW detection; diagnosis; identification; cytostatic; neuroprotective;  
KW neoplastic; immunomodulatory; relaxant; contraceptive; gynaecological;  
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;  
KW linkage analysis; tissue identification; tissue typing; forensic;  
KW neural; immune system; muscular; reproductive; gastrointestinal;  
KW pulmonary; cardiovascular; renal; proliferative.  
XX Homo sapiens.  
OS WO200055320-A1.  
XX 21-SEP-2000.  
XX 08-MAR-2000; 2000WO-US05989.  
XX 12-MAR-1999; 99US-0124270.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM;  
PI WPI; 2000-579444/54.  
XX N-PSDB; AAC98992.  
DR New nucleic acid that is a pancreatic cancer antigen for preventing,  
PT treating, or ameliorating a medical condition, particular pancreatic  
PT cancer, or for use in assays for diagnosing a pathological condition -  
XX Claim 11; Page 1115-1116; 1379pp; English.  
PS AAC98773 to AAC99231 encode the human pancreatic cancer associated

CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
CC neuroprotective, neoplastic, immunomodulatory, relaxant, contraceptive,  
CC gynaecological, cardiant and antiinflammatory activities, and can be used  
CC in gene therapy. The polynucleotide and proteins can be used for  
CC preventing, treating, or ameliorating a medical condition or in assays  
CC for diagnosing a pathological condition or a susceptibility to one in a  
CC subject. Binding partners to the proteins and the activity of the  
CC proteins can be identified. The pancreatic cancer antigens can be used to  
CC detect, treat or prevent pancreatic disorders, especially cancer.  
CC Agonists and antagonists to the antigens can be screened for. The  
CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
CC acid hybridisation probes that can be used in chromosome mapping, linkage  
CC analysis, tissue identification and/or typing and a variety of forensic  
CC and diagnostic methods. The proteins can be used to generate antibodies  
CC which are used to purify, detect and target the polypeptides, including  
CC both in vivo and in vitro diagnostic and therapeutic methods. The  
CC proteins can be used to treat or prevent neural, immune system, muscular,  
CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
CC sequences used in the exemplification of the present invention.

Sequence 395 AA;

Query Match 31.8%; Score 1737; DB 21; Length 395;  
Best Local Similarity 97.7%; Pred. No. 1.3e-135;  
Matches 335; Conservative 0; Mismatches 0; Indels 8; Gaps 1;  
QY 565 WFGSTVSKEDAMPEALKSLIFPNFEPHFKHTNFKLEIQRALWEGRSNAQIRDYORIG 624  
Db 21 wfgstvsksedampealkslifpnfepfhkhtnflkeieqrallwegrnaqirdyqrig 80  
QY 625 DYMLKNIQCMKHLAHLWKHSEALEALENGIKSSRLENFCRDFELQKCYLPLNTFLLR 684  
Db 81 dymlkniqcmkhlahlwkhsalealealengikssrlefcdrfelqkvcyplnflr 140  
QY 685 PLHRLMHYKQVLERLCKHHPPSHADFRDCRAALAEITEMVAQLHGTMIKMFNQKHLK 744  
Db 141 plhrlmhykqvierlckhphshadfrdcraalaeitemvaqlhgtmikenfkihelk 200  
QY 745 KDLIGDNLVVPGRFIRIGSLSKSGKGLQQRMFLENFNDVLLYTSRGLTASNQFVHGQ 804  
Db 201 kdligdnlvvpgrfirlgslsksgkglqqrmfllndvlltysrgltasngfkhvq 260  
QY 805 LPLVGMTESEDEWGVPHCLTLRGQRQSIIVAASRSEMEKVEDIQAIDLAEKSSSP 864  
Db 261 lplvgmtieesedewgvphcltlrgqrqsiivaasrsemekvvediqaiddlaeksssp 320  
QY 865 APEFLASSPPDNKSPDEATAADQESDDLSASRTSLERQAPHR 907  
Db 321 apeflassppdnkspdeataadqesddlsas-----phr 355

RESULT 4  
AAY91946  
ID AAY91946 standard; Protein; 1005 AA.  
XX AAY91946;  
AC AAY91946;  
XX 19-JUL-2000 (first entry)  
XX Human cytoskeleton associated protein 1 (CYSKP-1).  
DE Cytoskeleton associated protein; CYSKP-1; cancer; proliferative;  
KW autoimmunity; inflammatory; vesicle trafficking; neurological;  
KW cardiovascular; cell motility; reproductive; muscle disorder.  
XX Homo sapiens.  
OS Key Location/Qualifiers  
FH Modified-site 7  
FT /note= "potential phosphorylation site"

FT	Modified-site	15	"potential phosphorylation site"
FT	Modified-site	36	"potential glycosylation site"
FT	Modified-site	39	"potential phosphorylation site"
FT	Modified-site	58	"potential phosphorylation site"
FT	Modified-site	151	"potential phosphorylation site"
FT	Modified-site	159	"potential phosphorylation site"
FT	Modified-site	180	"potential phosphorylation site"
FT	Modified-site	194	"potential phosphorylation site"
FT	Modified-site	208	"potential phosphorylation site"
FT	Modified-site	212	"potential phosphorylation site"
FT	Modified-site	223	"potential phosphorylation site"
FT	Domain	225..263	"potential phosphorylation site"
FT	Domain	272..300	"signature sequence"
FT	Domain	321..367	"signature sequence"
FT	Modified-site	323	"signature sequence"
FT	Domain	379..408	"potential phosphorylation site"
FT	Modified-site	381	"signature sequence"
FT	Modified-site	402	"potential phosphorylation site"
FT	Domain	440..458	"potential phosphorylation site"
FT	Modified-site	449	"signature sequence"
FT	Modified-site	518	"potential phosphorylation site"
FT	Modified-site	530	"potential phosphorylation site"
FT	Modified-site	543	"potential phosphorylation site"
FT	Modified-site	544	"potential phosphorylation site"
FT	Modified-site	548	"potential phosphorylation site"
FT	Modified-site	614	"potential phosphorylation site"
FT	Modified-site	623	"potential phosphorylation site"
FT	Modified-site	647	"potential phosphorylation site"
FT	Modified-site	658	"potential phosphorylation site"
FT	Modified-site	673	"potential phosphorylation site"
FT	Modified-site	682	"potential phosphorylation site"
FT	Domain	718..721	"signature sequence"
FT	Modified-site	730	"signature sequence"
FT	Modified-site	744	"potential phosphorylation site"
FT	Modified-site	746	"potential phosphorylation site"
FT	Modified-site	748	"potential phosphorylation site"
FT	Modified-site	766	"potential phosphorylation site"

FT	Modified-site	/note=	"potential phosphorylation site"
FT	828	/note=	"potential phosphorylation site"
FT	854	/note=	"potential phosphorylation site"
FT	879	/note=	"potential phosphorylation site"
FT	884	/note=	"potential phosphorylation site"
FT	890	/note=	"potential phosphorylation site"
FT	944	/note=	"potential phosphorylation site"
FT	952	/note=	"potential phosphorylation site"
XX	W02000017355-A2.		
XX	30-MAR-2000.		
XX	17-SEP-1999;	99WO-US21565.	
XX	18-SEP-1998;	98US-0172226.	
XX	27-APR-1999;	99US-0131321.	
XX	(INCY-) INCYTE PHARM INC.		
XX	Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;		
PI	Guegler KJ, Patterson C, Azimzai Y, Baughn MR;		
XX			
XX	WPI: 2000-283582/24.		
DR	N-PSDB; AAA08581.		
XX			
PT	Human cytoskeleton associated proteins, used to treat cell		
PT	proliferative, autoimmune/inflammatory, vesicle trafficking,		
PT	neurological, cell motility, reproductive and muscle disorders		
XX			
PS	Disclosure; Page 79-82; 113pp; English.		
XX			
CC	AAV91946-61 show human cytoskeleton associated proteins 1 to 1		
CC	to CYSPK-16) respectively. The sequences can be used to treat		
CC	diagnose cancer and cell proliferative, autoimmune/inflammatory		
CC	trafficking, neurological, cardiovascular, cell motility, repara		
CC	and muscle disorders. Pharmaceutical compositions containing		
CC	CYSPK-16 can be used to treat or prevent disorders associated		
CC	decreased expression or activity of CYSPK (claimed), for exampl		
CC	atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis		
CC	cancers, autoimmune/antiinflammatory disorders such as allergi		
CC	asthma, acquired immunodeficiency syndrome (AIDS), Crohn's dis		
CC	diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scl		
CC	and trauma. CYSPK antagonists can be used to treat or prevent		
CC	disorder associated with increased expression or activity of C		
XX	(claimed).		
XX			
SQ	Sequence 1005 AA;		
	Query Match 13.2%; Score 720; DB 21; Length 1005;		
	Best Local Similarity 31.2%; Pred. No. 1.7e-50;		
	Matches 192; Conservative 102; Mismatches 222; Indels 100;		
Qy	34 TPSPGLVSIKIQMLDDTQAFVVPQAPCKVLLDVCNHLNLVGVDFGLFDPDKK		
Db	212 tktktvackvlllvgteyscdieknakqglvfdkvcchlnlliekdyfgllfqsp		
Qy	94 WLDLKLPIVKQIRRPKHVVVKKVAFPPDPHTQLOQELTRYFLFALQVKDLAGRLT		
Db	272 wldpakeikqrlnlpwift-fnvkfypppsqlteditryfclqlrldiasgrip		
Qy	154 TSAALLISHIVQSETGDPDEALDREH----LAKNKYIP-QQDALEDKYIEFFHNHIG		
Db	331 vthallqslvtqaelgdyd---peehgsldlsefqtaptqtkeleekvaelhkhtrg		

[illegible][illegible]

QY 209 AESDFOLLEIARRLEMYGIRLHPAKDRGTGKINIAVANTGILVFGQGTGKINAFNNAKVRK 268  
 Db 388 aqadsqflienakrismygdvllhahadsegvdikigvcangilkykdrlnrfawpklk 447  
 QY 269 LSEKRELIKLRPDANSAYODTLEFLMASRDFCKSFWKICVEHHAERFLFEKPKPKP 328  
 Db 448 isykrnsfykvrpaelqfestigfklpnhraakrlwkvcehntfrylv-speqppka 506  
 QY 329 VLFSGSFRSFGRTQKQVLDYVREGGKKVQFERKHSKIHSIRLASQPTLNSEVLEQ 388  
 Db 507 kfltlgskfrysgtrtaqtqastldrpaphfertskrvrs-rlsgapi---gvmddq 561  
 QY 389 SQQSTSLTFGBAGSPGQSGORKEPKVSGAGEGSHPSAPRSPAGNKQADGAASAP- 447  
 Db 562 slmk-----dfpagaag-----eisyagpgl-vslavvgdgdgrrevrptkaph 604  
 QY 448 -----TEEEVVVKORT-----QOSKPOPP- 467  
 Db 605 lqliegknslrvegdnlvyrhslmleeldkagedllkhqasiseikrnfmeestpepr 664  
 QY 468 -----QPSTGSLTGSPIH-----LSELSVNSQGVAPANVT-----LSPNLSP-DTK 507  
 Db 665 newekrritpislqtgssshetlniveekkraevgkderviteemngkeispgsggeir 724  
 QY 508 QASPLI-----SPLNDQACPTDDEGRRKRFTDKAYFIAKEVSTERTYTKDLEVIT 563  
 Db 725 kvepvtqkdstslsasseseeedvgeyrphhr---vtegtireeqeeyeeveep 781  
 QY 564 SWFQSTYSKEDAMPEA 579  
 Db 782 rpaakvvveeeavpea 797

RESULT 5

ID AAB95258 standard; Protein: 687 AA.  
 XX AAB95258;  
 AC AAB95258;  
 XX 26-JUN-2001 (first entry)  
 XX Human protein sequence SEQ ID NO:17435.  
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
 XX Homo sapiens.  
 XX EP1074617-A2.  
 XX 07-FEB-2001.  
 XX 28-JUL-2000; 2000EP-0116126.  
 XX 29-JUL-1999; 99JP-0248036.  
 XX 27-AUG-1999; 99JP-0300253.  
 XX 11-JAN-2000; 2000JP-0118776.  
 XX 02-MAY-2000; 2000JP-0183767.  
 XX 09-JUN-2000; 2000JP-0241899.  
 XX (HELI-) HELIX RES INST.  
 XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX WPI; 2001-318749/34.  
 XX Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs.  
 XX Claim 8; SEQ ID 17435; 2537pp + CD ROM; English.  
 PS  
 XX

CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX Sequence 687 AA;  
 SQ

Query Match 11.5%; Score 626; DB 22; Length 687;  
 Best Local Similarity 30.1%; Pred. No. 61e-43;  
 Matches 191; Conservative 92; Mismatches 221; Indels 130; Gaps 22;  
 QY 31 PPPTSGKLVSIKIOMLDDTQEAPEVQAPRQKVLDAVCNHLNLVEGDYFGLFPPDHK 90  
 Db 34 paagdsksiiatcrvslldgtadvldpkakagqelfqimyhldiesdyfqlrmdsaq 93  
 QY 91 ITWVLDLLKPIVQIRPKHVVVVFVFPFDPHTQLOBELTFLFALQVQDLAQGRLT 150  
 Db 94 vahldgtksikkqkigspcyhlrvkfyssennireeltryflvqlkqdsilsgkl 153  
 QY 151 CNDTSALLISHIVQSEIGDFDEALDREH---LAKNKYIP-QQDALEDKIVEFHHNHG 205  
 Db 154 cpfdtavqlaaynlgaeldydlia---ehspelvseifrvpiqteemelaifekwkyrg 210  
 QY 206 QTAESDFQLEIARRLEMYGIRLHPAKDRGTGKINIAVANTGILVFGQGTGKINAFNNAK 265  
 Db 211 qtpagaetnynlkakwlemygvdmhvkvkardgndyslgltpgtvlfegdkiglfwpk 270  
 QY 266 VRKLSFKRRKF-LIKLRPDANSAYOD-TLEFLMASRDFCKSFWKICVEHHAERFLFEK 322  
 Db 271 itrldfkknkltlvvvedddgkgehtfvfridhpkackhlwkavvehhaffrl-rgpv 329  
 QY 323 -KPKPKPVLFSGSFRSFGRTQKQVLDYVREGGKKVQFERKHSKIHSIRSL-----AS 376  
 Db 330 qkshrsrgfirlsrfrysgktyeqttktnk--arrstsferrpskrysrtilqmkacat 387  
 QY 377 QPTLNSEVLEQSQSTSLTFGBAGSPGQSGORKEPKVSGAGEGSHPSAPRSPAG 436  
 Db 388 kpeel-----svhnnvstqsgsqawgmrsalpsvpsis-----422  
 QY 437 NKQADGAASAPTEEEVEVVKDRTQOSKPOPPQ-----PSTGSLTSGPHSELSESVNS 487  
 Db 423 -----sapvpveie-----nlpspgtdqhdrcipnidlinspdilletti-- 464  
 QY 488 QGVVAPANVTLS-----PNLS-----PDTKQA-----SPLISPL 516  
 Db 465 -gdvigasdmetsqalndvnnvatrlpglgepeveyetlkdteeklkqlmensplispr 523  
 QY 517 LN-----DQACPRTDDEDEGRRR-KRFPTD-----KAYFIAKEVSTERTY 555  
 Db 524 snldvnninsqeevkvkleteclnnviespglnvrmvdpdfksnllkaqveavhkvtdsdl 583  
 QY 556 L--KDLEVITSWFQSTYSKEDAMP---EALKSLI 584  
 Db 584 lshknanvqdaatnsavlnennvplpksetlm 617

```

RESULT 6
AAW12522
ID AAW12522 standard; Protein; 913 AA.
XX AC AAW12522;
XX DT 22-MAY-1997 (first entry)
XX DE Protein tyrosine phosphatase that localises to focal adhesion.
XX KW Protein tyrosine phosphatase; PTPH1; focal adhesion;
XX KW protein tyrosine kinase; malignancy; cancer; gene therapy;
XX OS Homo sapiens.
XX FH Key
XX FT 30..357 Location/Qualifiers
FT FT /note= "N-terminal region shows homology with the
FT FT N-terminal regions of the talin family"
FT FT 357..663
FT FT /note= "central region includes sequences with
FT FT features of sites of phosphorylation by
FT FT casein kinase II and p34cdc2"
FT FT Modified-site 372
FT FT /label= Phosphorylation
FT FT /note= "putative p34cdc2 phosphorylation site"
FT FT Modified-site 381
FT FT /label= Phosphorylation
FT FT /note= "putative p34cdc2 phosphorylation site"
FT FT Modified-site 424..428
FT FT /label= Phosphorylation
FT FT /note= "putative casein kinase II phosphorylation
FT FT site"
FT FT Modified-site 438..442
FT FT /label= Phosphorylation
FT FT /note= "putative casein kinase II phosphorylation
FT FT site"
FT FT Modified-site 489..492
FT FT /label= Phosphorylation
FT FT /note= "putative casein kinase II phosphorylation
FT FT site"
FT FT Modified-site 514..517
FT FT /label= Phosphorylation
FT FT /note= "putative casein kinase II phosphorylation
FT FT site"
FT FT Modified-site 543..547
FT FT /label= Phosphorylation
FT FT /note= "putative casein kinase II phosphorylation
FT FT site"
FT FT Modified-site 607..610
FT FT /label= Phosphorylation
FT FT /note= "putative casein kinase II phosphorylation
FT FT site"
FT FT Region 604..931
FT FT /note= "C-terminal region shows homology to "
XX US5595911-A.
XX PN
XX DT 21-JAN-1997.
XX PD
XX PF 90US-0494036.
XX PR 01-MAR-1991;
XX PR 14-MAR-1990;
XX PR 16-AUG-1993;
XX PR 93US-0107420.
XX PA (COLD-) COLD SPRING HARBOR LAB.
XX PI
XX Tonks NK;
XX

```

```

DR WPI: 1997-107583/10.
XX N-PSDB; AAT58627.
PT DNA encoding protein tyrosine phosphatase - for gene therapy of
PT cancer
XX
XX PS Disclosure; Fig 1A-B; 12pp; English.
XX
CC A protein tyrosine phosphatase (PTPH1) (AAW12522) catalyses the
CC dephosphorylation of proteins in which tyrosyl residues have been
CC phosphorylated through the action of a protein tyrosine kinase
CC (PTK). It localises to focal adhesions, a major site of action of
CC oncogenic PTKs. PTPH1 is the product of a cDNA clone (AAT58627)
CC obt'd. from HeLa cells. The PTPH1 cDNA can be incorporated into a
CC vector (pref. retroviral) to allow expression of PTPH1 in mammalian
CC in sufficient quantities to overcome or counteract PTK activity.
CC Phosphorylation of tyrosine residues at abnormal levels is
CC prevented or reversed, resulting in the prevention or reversal of
CC malignancy of cells.
XX Sequence 913 AA;
SQ
Query Match 9.9%; Score 541; DB 18; Length 913;
Best Local Similarity 24.3%; Pred. No. 1.1e-35;
Matches 232; Conservative 150; Mismatches 346; Indels 228; Gaps 41;
QY 27 RGQKPPPTSGKLVSIKIQMLDDTQBAFEVPPQAPQKVLVDVNCNHLNLVEGDFGLFEP 86
DB 16 rtselepketrseivcsihfldgvvtkfvtkdgtgvlldvmhnlhgvtekyefglqhd 75
QY 87 DHKKITV-WDLLKPIVKQIRRPKHVVVVFVFFPDHTQLQEELTRYLFALQVQDLA 145
DB 76 ddsdvprwleaskrkqlkggfpcclhfrvrfipdpntlqgeqrthlyflqlkmdic 135
QY 146 QGRLTCTNDTSAALLISHIVQSEIGDFDEALDRE-HLAKNKYIPQDALED---KIVEFHH 201
DB 136 egritcplnsavvlasvayqshfgdynssihpgyysdshfipdq--edfltkveslhe 193
QY 202 NHIGQTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFOGFTINAF 261
DB 194 qhsqglkqseascyiniartldfygvelhgrdldhldmigiasagavayrkyictsfy 253
QY 262 NNAKVKLKFRRRLIKLRPDANSAYQDTLEFLMASRDCEKFWKICVHHAFRLFEE 321
DB 254 pwniikisfkrkffihgrkqaesrehivafnmnysrcknlwkscevhehtffga-kk 312
QY 322 PKPKPKPVL--FSRGSSFFSGRTOKQLVD-YVKE--GG-----HKKVQPERKHSKI 368
DB 313 llpqeknvlsqywtmgsr-----ntkksvnnqyckkvigmmvwnpamrrsisvehletk- 366
QY 369 HSTRSLASQPTELNSEVLEQSQGSTSLTFGEAGESPQSCRRGKPKVSAAGEPGSHSP 428
DB 367 -----slpsrspi-----tpnws-----prl-----rheir 389
QY 429 APRRSAGNKAQGAASAPT--EEEEVVKDRTOQSKPPQPPQSTGSLTSGPHLSUSVN 486
DB 390 kprhs-----sadnlanemtyitetedvfytykgslapg-dsdsevsqnrshqeslen 443
QY 487 SQGVAPANVTLSFNLSPDTKQASPLISPLLNDOACPRTDDEDEGRKRKRPDTPKAYIAK 546
DB 444 n-----padsyl-----tqsssvsspsnapgscspgvdqql-----lddfhrvtk 486
QY 547 EVSTTERTYLLKDLVITSWFQSTVSKEDAMPEALKSL-IPNFEPLHKTFTNLKETEQR 605
DB 487 ggstedas-----qyycdkndngdsylviriptded--gkfynlkgvvdqk 532
QY 606 LALWEGRSNAQ-----IRDYQRIQIGDVMLKNIQGMKH-----LAAHLKWHSEALEA 650
DB 533 mplvvsriropespadtcipklnegdgiivlingrdishethdqvvmfikasreshrel-a 591
QY 651 LENGIKSSRRLENFCRDFELQKVCYLPNTFLLRLPLRLMHYKQVLERLCKHHPPSHADF 710

```

Db 592 lvirravrsfadfkse delnql-----fpeaifpmc-----peggd- 628

Qy 711 RDCRAALAEITEMVAQLHGTMKME-----NFQKLHELKDLGIDNLVVPGRGF 760

Db 629 -----tlegsmaklgkgleagstvlqfeqlrykk-----pglai 662

Qy 761 IRLGSLKSLGKGLQRMFFLNDVLLY-TSRGLTASNQPKVHG-----QLPLYGWTIEE 814

Db 663 -----tfaklpqldknr-----ykdvlpydtttrvllqgnedyinasyyvnmelpaanlvnky 714

Qy 815 SEDEWGVPH-CL-----TLRGORQSIIVAASSRSEMEKWEVDIOWAIDLAEKSSSPAPEFL 869

Db 715 iatgdpiphtcaqfwvwdqklsily-----mtltitergrtkchqyw 758

Qy 870 ASSPPD-----NKSPDEATAADOESEDDLSASRTSLEROAPHRGNTMWHVCW 916

Db 759 -pdppdvmmhbgfhicqsedctiayvsremlvntqtgehtvth-----lqyvaw 809

RESULT 7

AA25156

ID AAY25156 standard; Protein; 913 AA.

XX AC AAY25156;

XX DT 07-SEP-1999 (first entry)

XX DE Human PTPH1 protein.

XX KW PTPH1; human; protein tyrosine phosphatase; focal adhesion; cancer;

KW localisation; treatment; overexpression; oncogenic; cell transformation;

KW prevention; phosphotyrosine; disease; malignant.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Region 30..337

FT /note= "region of homology to the N-terminal domain

FT of band 4.1, ezrin and talin. This region is

FT known to be important for localisation to focal

FT adhesions"

FT Modified-site 372

FT /note= "potential p34cdc2 phosphorylation site"

FT Modified-site 381

FT /note= "potential p34cdc2 phosphorylation site"

FT Modified-site 424..428

FT /note= "potential casein kinase II phosphorylation site"

FT Modified-site 434

FT /note= "potential p34cdc2 phosphorylation site"

FT Modified-site 438..442

FT /note= "potential casein kinase II phosphorylation site"

FT Modified-site 489..492

FT /note= "potential casein kinase II phosphorylation site"

FT Modified-site 514..518

FT /note= "potential casein kinase II phosphorylation site"

FT Modified-site 543..547

FT /note= "potential casein kinase II phosphorylation site"

FT Modified-site 607..610

FT /note= "potential casein kinase II phosphorylation site"

XX US5863781-A.

XX PD 26-JAN-1999.

XX PF 04-DEC-1996; 96US-0759536.

XX PR 01-MAR-1991; 91US-0663579.

PR 14-MAR-1990; 90US-0494036.

PR 16-AUG-1993; 93US-0107420.

XX 04-DEC-1996; 96US-0759536.

PA (COLD-) COLD SPRING HARBOR LAB.

XX

PI Tonks NK;

XX WPI; 1999-131308/11.

DR N-PSDB; AAX78463.

XX Protein tyrosine phosphatase PTPH1 - encoded by DNA of HeLa cells

PT Claim 1; Fig 1A-B; 12pp; English.

XX This sequence represents a novel protein tyrosine phosphatase, PTPH1,

CC isolated from HeLa cells. The protein of the invention appears to

CC localise to focal adhesions and is therefore potentially useful in the

CC treatment of cancer. Overexpression of PTPH1 can be used to counter the

CC effects of oncogenic protein tyrosine kinases such as those of

CC transforming viruses and for interfering with or reversing cell

CC transformation. This would provide a means of preventing or reversing

CC abnormally high levels of phosphotyrosine associated with any disease or

CC condition such as preventing or reversing malignancy associated with the

CC activity of a protein tyrosine kinase.

XX Sequence 913 AA;

SQ

Query Match 9.9%; Score 541; DB 20; Length 913;

Best Local Similarity 24.3%; Pred. No. 1.1e-35;

Matches 232; Conservative 150; Mismatches 346; Indels 228; Gaps 41;

Qy 27 RGOKPPPTPSGKLVSIKIOMLDDTQEAPEVPORAPGKVLDDAVCNHNLNVEGDYFGLFEP 86

Db 16 rtseipkekrsevischfhdgvgvqtkgtqdgvgldmvmhnhlgvtekeyfiglhd 75

Qy 87 DHKKITV-WLDLLKPIVKQIRRPKHVVVWFVAFPPDPHTQLQOEELTRYLFALQVKODLA 145

Db 76 ddsavdsprwleaskpirkqlkggfpctlhfrvrfpdpntlqqeqtrhlyfqlkmdic 135

Qy 146 QGRLTNDTSAALLSHIVOSEIGDPEALDRE-HLAKNYIFQOQDALED---KIVFHH 201

Db 136 egrltcplnsavvlasyavqshfgydnssihhpgylsdshfipdn--edfltkveslhe 193

Qy 202 NHGQTPAESDFQLLETARLEMYGIRLHPAKDREGTKINLAVANTGILYFQGTKNF 261

Db 194 qhsglkqasecyiniartldfygvelhsgrdlnldlmglasagvayrvkyictsfy 253

Qy 262 NNAVKRLSKFRKRLTKLRPDANSAYQDTLEFLMASRDECFKSWKICVHHAFRFRLEE 321

Db 254 pwnilkisfrkffihqkqaesrehivafnmlyrsknlwkcscvehhtffga-kk 312

Qy 322 PKPKPKPVL---FSRGSSFRFSGRTQKVLD-VYKE---GG-----HKKVQPERKHSKI 368

Db 313 llpqeknvlsgywtmgsr-----ntkksvnnqyckkviggmwnpanrrrsalsvehletk- 366

Qy 369 HSIRSLASQPTELNSEVLEOSQOSTSLTFGEAESGSGSCRRGKPKVSGAGEPGHSP 428

Db 367 -----slpsrppli-----tpnwr-----prl-----rheir 389

Qy 429 APRRSPAGNKQADGAASAPT--EEEEVVKDRTQQSKPPQPPSTGSLTSGPHSEL SVN 486

Db 390 kprhs-----sadnlanemyitetedvfytqkslapq-dsdsevsqnrspqhesl sen 443

Qy 487 SQGGVAPANVTLSFNLSPDTKQASPLISLLNDOACPRTDDEDEGRKRKRPPTKAYIAK 546

Db 444 n-----paqsy l-----tqksssvspssnagpcspgdvqdql-----lddfhrvtk 486

Qy 547 EVSTTERTYTKDLEVTISWFSQTVSKEDAMPEALKSI-IPFNPEPLKHTNFKLETEOR 605

Db 487 ggstedas-----qyycdkndngdsylviriltpded--gkfgfnlkggvdqk 532

Qy 606 LALWEGRSNAQ-----IRDYQIRIGDVMLKNIQGMKH-----LAAHLWKHSEALEA 650

Db 533 mplvvsrinpespadtcipklnegdqivlingdisethdgvvmfikasreshrel-a 591

Qy 651 LENGIKSSRRLENFCRDFELQKVCYLPNTFTLLRLPLRLMHYKQVLERLCKHPPPSHADF 710









QY 501 NLSPTTKOASPL-ISPLNDQACPT 525  
 Db 600 hhsvqtfdedslpvaahsleqvsepl 625

RESULT 12  
 AAY34158  
 ID AAY34158 standard; Protein; 1174 AA.  
 AC AAY34158;  
 DT 09-NOV-1999 (first entry)  
 DE Human protein tyrosine phosphatase, PTP-D1, protein sequence.  
 XX PTP-D1; human; protein tyrosine phosphatase; neoplastic formation;  
 KW PTPase; growth factor signal transduction; cell cycle progression;  
 KW cancer; diabetes; cellular phosphotyrosine metabolism.  
 XX Homo sapiens.  
 OS  
 XX US5955592-A.  
 PN  
 XX 21-SEP-1999.  
 XX 27-APR-1994; 94US-0234440.  
 XX 27-APR-1994; 94US-0234440.  
 PR 03-AUG-1992; 92US-0923740.  
 XX  
 XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 PA  
 XX Moller KB, Moller NPH, Ullrich A;  
 PI  
 XX WPI; 1999-539599/45.  
 DR N-PSDB; AA211221.  
 XX  
 XX Nucleic acids encoding protein tyrosine phosphatase PTP-D1  
 PT useful for identification of PTP-D1 modulators  
 XX  
 XX Claim 1; Fig 5; 63pp; English.  
 XX  
 CC This sequence is the protein tyrosine phosphatase of the invention,  
 CC designated PTP-D1. The new PTP proteins are a subfamily of protein  
 CC tyrosine phosphatases (PTPases) involved with growth factor signal  
 CC transduction, cell cycle progression, and neoplastic formation. The DNA  
 CC sequence may be used in the recombinant production of PTP-D1 according to  
 CC standard DNA methodology. The protein expressed may then be used in  
 CC assays to identify modulators of its enzymatic activity and in the  
 CC production of antibodies. The nucleic acids may also be used in assays to  
 CC detect and quantify expression of PTP-D1 in samples. Methods identifying  
 CC normal or mutant PTP-D genes, or for measuring amount or activity of the  
 CC protein can serve as methods for identifying susceptibility to cancer,  
 CC diabetes, or other disorders associated with alterations in cellular  
 CC phosphotyrosine metabolism.  
 XX  
 SQ Sequence 1174 AA;

Query Match 7.2%; Score 396; DB 20; Length 1174;  
 Best Local Similarity 22.5%; Pred. No. 1.8e-23;  
 Matches 141; Conservative 103; Mismatches 248; Indels 134; Gaps 18;

QY 25 LERGOPTTPSKLYSIKIMLDLDTQAEFVQAPAGKVLDAVCNHLNVEGDYFGL 84  
 Db 9 lkrtrvtsksclva-riqlinnfevftlsvestqgesleavagrlrevtyfslw 67  
 QY 85 FPDHKITVWLDLLKPIVKQI-RRPKHVVKVFFPPDHTQLQELTRYLFALQVKOD 143  
 Db 68 ynnkqqrwvdlkplkqldkyleptvfygvfypsvsqleqitryqyqlkld 127  
 QY 144 LAOGRUTCNTSAALLISHIVQSEIGDFDREHLAKNKYIP-----QQDALED---K 195

Db 128 ilegspctleqaiqlaglavqadfgdqesqdfkqlfapvqglvdekvleatqk 187  
 QY 196 IVEFHNNHIGTTPAESDFQLLEIARLEMYGIRLPAKREGTKINLAVANTGILV-FQG 254  
 Db 188 vallhkyrgltapdaemlymqevermdgygeesypakdsqgsdisigaclegifvkhn 247  
 QY 255 FTKINAFNAKVRKLSFKRRKFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHA 314  
 Db 248 grhpvvrwhdianmhsksfalel---ank--eciqfqtmedmetakyiwrivvarhk 302  
 QY 315 FFLRLE-----EPKPKPKPVLFSGSSFRFSGR-----TQKQVL 348  
 Db 303 fyrlncnliqtvtvnprrrrssrmslpkqpymppppqlhynghtyepvassqdn 362  
 QY 349 DYVKEG---HKVQFER-----KHSKIHSIRS-----LASOPTEL 381  
 Db 363 fvpngngyysqtsidraqidfngrirngsvysahstnlnnpqylqpsmssnpsit 422  
 QY 382 NSEVL-----EQSQSTSLTFG----- 398  
 Db 423 gsdvmpdyplshrsavippsyrptdyetvmkqlnrglvhaergshlninissya 482  
 QY 399 -----EQAESPQGQSCRCGKEPKVSAEGSPHSPAPRRSPAGNKQADGA 443  
 Db 483 ysrrpaalvysqpeirehaqlpsaaahcftlsysfhspsypypaerppvavsvpel 542  
 QY 444 ASAPTEEEVEVKD--RTQOSKRPQPPQSTGSLTSGPHLS-ELSVNSQGVAPANTVLS 500  
 Db 543 tnaqlgaqdypspnlmrtqvyrrpppypprpanstpdlsrhllyssn---pdlltrrv 599  
 QY 501 NLSPTTKOASPL-ISPLNDQACPT 525  
 Db 600 hhsvqtfdedslpvaahsleqvsepl 625

RESULT 13  
 AAB53356  
 ID AAB53356 standard; Protein; 635 AA.  
 XX  
 AC AAB53356;  
 XX  
 DT 09-MAR-2001 (first entry)  
 DE Human colon cancer antigen protein sequence SEQ ID NO:896.  
 XX  
 KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;  
 KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
 KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;  
 KW neural disorder; immune system disorder; muscular disorder;  
 KW reproductive disorder; gastrointestinal disorder; renal disorder;  
 KW infectious disease; cardiovascular disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200005351-A1.  
 XX  
 PD 21-SEP-2000.  
 XX  
 PF 08-MAR-2000; 2000WO-US05883.  
 XX  
 PR 12-MAR-1999; 99US-0124270.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX Rosen CA, Ruben SM;  
 PI  
 XX WPI; 2000-587534/55.  
 DR N-PSDB; AAC98113.  
 XX  
 PT Colon cancer associated gene sequences, referred to as colon cancer  
 PT antigens, useful for the treatment, prevention, and diagnosis of colon

disorders such as colon cancer -

Claim 11; Page 1449-1451; 2104pp; English.

AAC97991 to AAC98763 encode the human colon cancer associated proteins, called human colon cancer antigens, given in AAB53234 to AAB54006. The human colon cancer antigens can have cytostatic, cardioactive, muscular, neuroprotective, immunomodulatory, gynaecological, gastrointestinal, vulnary, nephrotropic, antineoplastic and antibacterial activities, and can be used in gene therapy. The colon cancer antigen polynucleotides, proteins and antibodies to the proteins are useful for the prevention, treatment and diagnosis of colon disorders, such as colon cancer. The polynucleotides may be used in diagnostics and research, such as for chromosome identification, and as hybridisation probes. The proteins may also be used to prevent diseases such as neural disorders, immune system disorders, muscular disorders, reproductive disorders, gastrointestinal disorders, wounds, renal disorders, infectious diseases, and cardiovascular disorders. AAC98764 to AAC98772 and AAC98773 represent sequences used in the exemplification of the present invention.

Sequence 635 AA;

Query Match 6.7%; Score 367.5; DB 21; Length 635;  
Best Local Similarity 24.0%; Pred. No. 1.6e-21;  
Matches 125; Conservative 98; Mismatches 226; Indels 71; Gaps 16;

QY 7 RPTGSRGAPEN-----SGISTLERGOKPPTPTSGKLVSIKIQMLDDTQAEAFVQR 59  
DB 19 RSTAGAGLGRCCWVVFSGFSGHQPMP-----KPINVVTMDAAE-FAIQPN 72  
QY 60 APGKVLDDAVCNHLNVEGDYFGLFEPDHPKKTITVWDLKPI-VKQIRPKHVVVVKV 118  
DB 73 TTKGQIDQVVKTVIGREVWYFGLHYDNGKFTWIKDKKSAQVKNPQKFRK 132  
QY 119 FFPDP-HTQLEBELRYLALQVQDLAQGRVTCNDTSALLSHIVQSEIGDFDALDR 177  
DB 133 FYPEDVAEELIGDTKQIFLQVKEGILSDIYCPBPETAVLGSYAVQAKFGDYKVEKH 192  
QY 178 E-HLAKNKYIPQ-----QDALEDKIVERHNNHIGOTPAESDFQLEIARRLEMYGI 227  
DB 193 SGYISSEIRLIPQVMDGHKLTDRQWEDRIGVWHAHRGMLKDNAMLEYKIQADLEMYGI 252  
QY 228 RLHPAKDREGTKINLAVANTGILVFGQTKIN---AFNNAKVKLSFKRRKRLI----KL 280  
DB 253 NYFEIKNKGTDLWGLDGLNLIEYKDDKLTQKIGFVSEIRNISFNDDKFKVQPIDKK 312  
QY 281 RPDANSAYQDTLEFLMASRDFCSFWKICVEHHAFFRLPEEPKPKPKPVLFSGSSFRFS 340  
DB 313 APD-----FVYAPRLRINKRIQLCMGNHLYMRRIRKPDITIEVQGMKAQAREKHQ 364  
QY 341 GRTOKQVLDVYKE-----GGHKKVQFERKHSKTHSRSLASOPTLENSVLEQSQSTSL 395  
DB 365 KQIERGQLEKRRRETVEREKEGMREKEELMLRIQDYEEKCKKAERLSEIQIRALQL 424  
QY 396 -----TFEGAE--SPGGQSCRRGKE-----PKVSAGEPGSHSPARRSPAG 436  
DB 425 EEEKRAQEAERLEADRMAALRAKEELERQAVDQIKSQEQLAAELAEYAKIALLEEER 484  
QY 437 NKQADGAAS---APTDEEEVVKDRTO-----QSKPPQPPQ 469  
DB 485 RREDEVEEQHRAKEAQDVLVTKKEELHLVMTAPPPIPP 524

RESULT 14  
AAY27443  
ID AAY27443 standard; protein; 586 AA.  
XX  
AC AAY27443;  
XX  
XX 26-NOV-1999 (first entry)  
XX

Amino acid sequence of human ezrin polypeptide.

Pharmaceutical; ezrin; mutant; tumor; metastasis; human.

Homo sapiens.

Key Location/Qualifiers  
Misc-difference 354 /note= "the Tyr at this position can be mutated (preferably to a Phe) to construct an ezrin mutant of the invention"

WO9947150-A2.  
23-SEP-1999.  
18-MAR-1999; 99WO-EP02054.  
18-MAR-1998; 98US-0040725.  
(CURI-) INST CURIE.  
(CNRS ) CNRS CENT NAT RECH SCI.  
Arpin M, Crepaldi T, Gautreau A, Louvard D;  
WPI; 1999-561851/47.  
New composition for prevention and treatment of tumors and metastasis  
Example 1; Fig 1; 31pp; English.  
The invention provides a pharmaceutical composition containing ezrin protein, RNA or DNA mutated on tyrosine 353, or a functional fragment or derivative of the ezrin mutant. The new composition is useful for prevention and/or treatment of tumors, and especially metastasis. The present sequence represents the amino acid sequence of human ezrin (before the maturation by deletion of the first amino acid Met).  
Sequence 586 AA;

Query Match 6.5%; Score 356.5; DB 20; Length 586;  
Best Local Similarity 23.9%; Pred. No. 1.1e-20;  
Matches 115; Conservative 96; Mismatches 212; Indels 59; Gaps 14;

QY 38 KLVSIKIQMLDDTQAEAFVQPARAPGVLLDAVCNHLNVEGDYFGLFEPDHPKKTITVWDL 97  
DB 3 KPINVVTMDAAE-FAIQNTTGKQIFDQVVKTVIGREVWYFGLHYDNGKFTWIKL 61  
QY 98 LKPI-VKQIRPKHVVVVKVVFPPDP-HTQLEBELRYLALQVQDLAQGRVTCNDTS 155  
DB 62 DKKVSAQVKNPQKFRKFPEDVAEELIGDTKQIFLQVKEGILSDIYCPPET 121  
QY 156 AALLSHIVQSEIGDFDEALDRE-HLAKNKYIPQ-----QDALEDKIVERHNNHIG 205  
DB 122 AVILGSYAVQAKFGDYKVEKHSGYISSEIRLIPQVMDGHKLTDRQWEDRIGVWHAHRG 181  
QY 206 QTPAESDFQLEIARRLEMYGIRLHPAKDREGTKINLAVANTGILVFGQTKIN---AFN 262  
DB 182 MLKDNAMLEYKIQADLEMYGINYFEIKNKGTDLWGLDGLNLIEYKDDKLTQKIGFP 241  
QY 263 WAKVRKLSFKRRKRLI----KLRPDANSAYQDTLEFLMASRDFCSFWKICVEHHAFFRL 318  
DB 242 WSEIRNISFNDDKFKVQPIDKKAPD-----FVYAPRLRINKRIQLCMGNHLYM 293  
QY 319 FEEPKPKPKPVLFSGSSFRFSRGTOKQVLDYVKE-----GGHKKVQFERKHSKHSIRS 373  
DB 294 RRPDITIEVQGMKAQAREKHQKQIERGQLEKRRRETVEREKEGMREKEELMLRIQD 353  
QY 374 LASQPTLENSVLEQSQSTSL-----TFEGAE--SPGGQSCRRGKE----- 414  
DB 354 YEEKTKAERLSEIQIRALQEEERKRAQEAERLEADRMAALRAKEELERQAVDQIKS 413

QY 415 PKVAGEPGSHPSAPRRSPAGNKQADGAAS---APTEEEVEVVKORTQ-----QSKPQPP 467  
 Db 414 qeqlaaelaeytakialleearrrkedeveeqhrrakeagddlvktkeelhlymtapppp 473  
 QY 468 QP 469  
 Db 474 pp 475

RESULT 15  
 AAM25892  
 ID AAM25892 standard; Protein; 305 AA.  
 AC AAM25892;  
 DT 16-OCT-2001 (first entry)  
 DE Human protein sequence SEQ ID NO:1407.  
 KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
 KW anti-inflammatory; antirheumatic; antiarthritic; immunosuppressive;  
 KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
 KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
 KW antiaggregant; haemostatic; vulnerary; antiulcer; osteopathic; eczema;  
 KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
 KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
 KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
 KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
 KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmune;  
 KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
 KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
 KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
 KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
 KW neurological disorder.  
 OS Homo sapiens.  
 XX WO200153455-A2.  
 PN 26-JUL-2001.  
 PD 22-DEC-2000; 2000WO-US35017.  
 PF 23-DEC-1999; 99US-0471275.  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 XX (HYSE-) HYSEQ INC.  
 PA Tang YT, Liu C, Drmanac RT;  
 PI WPI; 2001-457603/49.  
 XX N-PSDB; AAH99833.  
 DR Isolated human polynucleotides encoding polypeptides, useful for the  
 PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -  
 PT Claim 20; Page 287; 1217pp; English.  
 XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to  
 CC AAM25963. The proteins can have activities based on the tissues and  
 CC cells they are expressed in, such as: anti-inflammatory; antirheumatic;  
 CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
 CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
 CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnerary;  
 CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;  
 CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
 CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
 CC encoding them can be used in gene therapy, antisense therapy and vaccine  
 CC production. The proteins and polynucleotides are useful for screening for  
 CC agonists or antagonists of a protein and for the treatment and diagnosis  
 CC of disorders associated with the activity of a protein e.g. inflammation,

CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
 CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
 CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
 CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
 CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
 CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
 CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
 CC neurological disorders.  
 XX  
 SQ Sequence 305 AA;

Query Match 6.3%; Score 342.5; DB 22; Length 305;  
 Best Local Similarity 33.1%; Pred. No. 5.9e-20;  
 Matches 98; Conservative 48; Mismatches 109; Indels 41; Gaps 13;

QY 166 SEIGDFDEALDREH---LAKNKYIP-QQDALEDKIVEFHNNHIGQTPAESDFQLLEIAR 220  
 Db 8 aelgdydla---ehspelysefvpqlteemelaifekwkeyrggtapaetnylnkak 64  
 QY 221 RLEMYGIRLHPAKDREGTKINLAVANTGIIIVFOGFTKINAFNWKVKLSFKKRF-LIK 279  
 Db 65 wlemygvdmhvkarogndysigltptgvivfegdkiglfwpktrldfknkktlvv 124  
 QY 280 LRPDANSAYQD-TLEFLMASRDFCKSFWKTCVEHHFAFFRLFEFP--KPKPKPVLFSGSS 336  
 Db 125 vedddgqkeqhtfvrlidhpkackhlwkcavehhaffrl-rgpvqksshrgsfirlgsr 183  
 QY 337 FRFSGRTQQLVDYVVEGGHKKVQFERKSKIHISRL-----ASQTELNSEVLQSQQ 391  
 Db 184 frysgkteygttktnk--arrstsferrpskrysrtrtlgmakatkepel-----svh 234  
 QY 392 STSLTFGEAESPGGOSCRRGKPKVSAGEPGSHSPAP-----RRSPAGNKQAD 441  
 Db 235 nnvstqsgsqawgmrsalpvpspsiss-----apvpveienlpqsp-gtdqhd 282

Search completed: December 6, 2001, 08:51:40  
 Job time: 332 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2001, 08:37:48 ; Search time 4659.49 Seconds  
(without alignments)  
12186.589 Million cell updates/sec

Title: US-09-555-342A-1  
Perfect score: 3442  
Sequence: 1 ccccgagccgcgcgcgtg.....agtattaaacattgtcatt 3442

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: gb\_ba:\*
- 2: gb\_hg:\*
- 3: gb\_in:\*
- 4: gb\_om:\*
- 5: gb\_ov:\*
- 6: gb\_pat:\*
- 7: gb\_ph:\*
- 8: gb\_pl:\*
- 9: gb\_pr:\*
- 10: gb\_ro:\*
- 11: gb\_sts:\*
- 12: gb\_sy:\*
- 13: gb\_un:\*
- 14: gb\_vi:\*
- 15: em\_ba:\*
- 16: em\_fun:\*
- 17: em\_hum:\*
- 18: em\_in:\*
- 19: em\_om:\*
- 20: em\_or:\*
- 21: em\_ov:\*
- 22: em\_pat:\*
- 23: em\_ph:\*
- 24: em\_pl:\*
- 25: em\_ro:\*
- 26: em\_sts:\*
- 27: em\_sy:\*
- 28: em\_un:\*
- 29: em\_vi:\*
- 30: em\_htgo\_hum:\*
- 31: em\_htgo\_inv:\*
- 32: em\_htgo\_rod:\*
- 33: em\_htg\_hum:\*
- 34: em\_htg\_inv:\*
- 35: em\_htg\_rod:\*
- 36: em\_htg\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

8

Result No.	Score	Query Match	Length	DB ID	Description
1	3442	100.0	3442	9	AB008430 Homo sapi
2	1054	30.6	2632	10	BC004009 Mus muscu
3	1037.4	30.1	3997	9	AB018336 Homo sapi
4	1019.2	29.6	3719	10	BC009153 Mus muscu
5	342.8	10.0	2888	9	AL161984 Homo sapi
6	339.2	9.9	181800	2	AL137249 Homo sapi
7	322.2	9.4	483	11	G22662 Human STS W
8	286.6	8.3	1438	9	AL122052 Homo sapi
9	254	7.4	96183	9	AL161896 Human DNA
10	254	7.4	152053	2	AC022669 Homo sapi
11	199.2	5.8	170916	9	AL136300 Human DNA
12	194	5.6	2595	9	AF156225 Homo sapi
13	193.4	5.6	152053	2	AC022669 Homo sapi
14	192.4	5.6	2867	9	HMELIA Human struc
15	192.4	5.6	3043	9	HUMEMP41 Human
16	189.4	5.5	5942	10	MUS41SP Mus musculu
17	187	5.4	216	11	G23578 human STS W
18	187	5.4	2882	10	AF106702 Mus muscu
19	185.2	5.4	6102	10	AB019256 Rattus no
20	183.6	5.3	5855	10	AB019257 Rattus no
21	181.8	5.3	3619	9	BC010674 Homo sapi
22	181.8	5.3	3643	9	M68941 Human prote
23	181.2	5.3	2590	4	AF222767 Bos tauru
24	181.2	5.3	3661	10	AB032828 Rattus no
25	181.2	5.3	4543	10	AB032827 Rattus no
26	180.4	5.2	6263	9	AB002336 Human mRN
27	179.6	5.2	5921	3	L27467 Drosophila
28	176.6	5.1	2964	10	AF044312 Mus muscu
29	174	5.1	3320	10	AF061283 Mus muscu
30	172.4	5.0	2758	5	M20621 X.laevis cy
31	169.6	4.9	4336	9	AF027299 Homo sapi
32	166.8	4.8	3451	10	AB032366 Mus muscu
33	165	4.8	4051	10	AF152247 Mus muscu
34	162.4	4.7	393	11	G60059 SHGC-130885
35	161.8	4.7	2717	10	AF177146 Mus muscu
36	161.6	4.7	3309	9	AF069072 Homo sapi
37	161.6	4.7	3387	9	BC006141 Homo sapi
38	161.6	4.7	4446	9	AB023204 Homo sapi
39	157.6	4.6	1973	9	AB070172 Macaca fa
40	154.2	4.5	2677	9	HMELI Human eryth
41	154	4.5	42014	2	AC018327 Drosophil
42	154	4.5	160710	3	AC009537 Drosophil
43	154	4.5	164713	3	AC007589 Drosophil
44	154	4.5	303209	3	AE003604 Drosophil
45	144	4.2	3984	9	M64572 Human prote

## ALIGNMENTS

RESULT 1

AB008430	AB008430	3442 bp	mrna	PRI	13-FEB-1999
LOCUS	Homo sapiens	3442 bp	complete cds.		
DEFINITION	AB008430				
ACCESSION	AB008430				
VERSION	AB008430.1	GI:2766164			
KEYWORDS	CDEP.				
SOURCE	Homo sapiens	embryo cartilage chondrocyte	cdna to mrna.		
ORGANISM	Homo sapiens				
REFERENCE	1	(bases 1 to 3442)			
AUTHORS	Koyano, Y., Kawamoto, T. and Kato, Y.				
TITLE	Submitted (22-OCT-1997)				
JOURNAL	Kawamoto, Hiroshima University School of Dentistry, Department of Biochemistry, 1-2-3 Kasumi Minami-ku, Hiroshima, Hiroshima 734, Japan (E-mail: tkawamo@ipc.hiroshima-u.ac.jp, Tel: 082-257-5688, Fax: 082-257-5629)				
REFERENCE	2	(sites)			

AUTHORS Koyano,Y., Kawamoto,T., Shen,M., Yan,W., Noshiro,M., Fujii,K. and Kato,Y.  
 TITLE Molecular cloning and characterization of CDEP, a novel human protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors  
 JOURNAL Biochem. Biophys. Res. Commun. 241 (2), 369-375 (1997)  
 MEDLINE 98086358  
 FEATURES Location/Qualifiers  
 1..3442  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_type="chondrocyte"  
 /dev\_stage="embryo"  
 /tissue\_type="cartilage"  
 49..3186  
 /function="rho Guanine Nucleotide Exchange Factor"  
 /note="Band 4.1 superfamily"  
 /codon\_start=1  
 /product="CDEP"  
 /protein\_id="BAA24267.1"  
 /db\_xref="GI:2766165"  
 /translation="MGEIEQRPFGSRILGAPENSGISTLERGQRPPTPSKLVSIKI QMLDDTQAFVQRPAGPKVLLDVAENHLNLYEGDYFGLEPDHKKITVNDLLKPIV KQIRPKHVVFVYKFPDPDHTQLEELTRYFALQVKQDLAQRCTNDSALALI SHIVQSEIGDFDEALDREHLAKNKYIPOODALEDKIVFEHFNHIGQTPAESDFOLLEI ARRLMYGIRLHPAKDRGTINLAVANTGILVFGQFTKINAFNWKVRKLSFKRRF LILKRPDANSAYQDTLEFLMSRDECFKFWKICVEHAFPLFEPRPKPKPVPLESRG SSFPGSGTQKQVLDYVREGGKPKVQFERKSKHSIRLSAQSPTELNSEVLEQSQQS TSLTFGEAGSGPGSGSKEPKVSGAGEPSHPSPAPRSPAGNQKQDGAASAPTEE EEEVVDKTOQSKPQPPQSTGSLTSPHLSVNSQGVAPANVTLSPLNSPDTKQ ASPLISPLNDQACPTDEDEGRKRPPTDKAYFIAKEVSTERTYLLKDLLEVITSWF QSTVSKEDAMPEALKSLFPNPEPLKFTNFKLEIQRALWEGRSNAQIRDXQRIG DMLKNIGMKHLAHLKHSLEALEENGKISSRLENFCRDFELQKVCYLPNLTFL LRPLRLHMYKQVLERLCKHHPSHADPROCRALAEITEMVAQLHGMWIKMENFQKL HELKDLGIDNLYVVPREFIRLGSLSKLGKIQORMFFLNFNDVLLYTSRGLTASNQ FKVHGQPLVMTTESEDEWGVPHCLTLRGQRQSIIIVASSRSEMEXWVEDIQMAID LAEKSSPAPFELASSPPDNKSPDEATAQDESDLSASRTSLERQAPHRGNTMVHV CWHRTNSVYDFSTAVENOLSGNLLRKFKNMGWOKLWVFTNQLFFYKSHODNHP LASPLPLGYSLTIPSESENIQKDYVFKLHPKSHVYFPAESEYTFERMEVIRATSS ASRPHVSHKESLVY"

## FEATURES

## Source

## CDS

1..3442  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_type="chondrocyte"  
 /dev\_stage="embryo"  
 /tissue\_type="cartilage"  
 49..3186  
 /function="rho Guanine Nucleotide Exchange Factor"  
 /note="Band 4.1 superfamily"  
 /codon\_start=1  
 /product="CDEP"  
 /protein\_id="BAA24267.1"  
 /db\_xref="GI:2766165"  
 /translation="MGEIEQRPFGSRILGAPENSGISTLERGQRPPTPSKLVSIKI QMLDDTQAFVQRPAGPKVLLDVAENHLNLYEGDYFGLEPDHKKITVNDLLKPIV KQIRPKHVVFVYKFPDPDHTQLEELTRYFALQVKQDLAQRCTNDSALALI SHIVQSEIGDFDEALDREHLAKNKYIPOODALEDKIVFEHFNHIGQTPAESDFOLLEI ARRLMYGIRLHPAKDRGTINLAVANTGILVFGQFTKINAFNWKVRKLSFKRRF LILKRPDANSAYQDTLEFLMSRDECFKFWKICVEHAFPLFEPRPKPKPVPLESRG SSFPGSGTQKQVLDYVREGGKPKVQFERKSKHSIRLSAQSPTELNSEVLEQSQQS TSLTFGEAGSGPGSGSKEPKVSGAGEPSHPSPAPRSPAGNQKQDGAASAPTEE EEEVVDKTOQSKPQPPQSTGSLTSPHLSVNSQGVAPANVTLSPLNSPDTKQ ASPLISPLNDQACPTDEDEGRKRPPTDKAYFIAKEVSTERTYLLKDLLEVITSWF QSTVSKEDAMPEALKSLFPNPEPLKFTNFKLEIQRALWEGRSNAQIRDXQRIG DMLKNIGMKHLAHLKHSLEALEENGKISSRLENFCRDFELQKVCYLPNLTFL LRPLRLHMYKQVLERLCKHHPSHADPROCRALAEITEMVAQLHGMWIKMENFQKL HELKDLGIDNLYVVPREFIRLGSLSKLGKIQORMFFLNFNDVLLYTSRGLTASNQ FKVHGQPLVMTTESEDEWGVPHCLTLRGQRQSIIIVASSRSEMEXWVEDIQMAID LAEKSSPAPFELASSPPDNKSPDEATAQDESDLSASRTSLERQAPHRGNTMVHV CWHRTNSVYDFSTAVENOLSGNLLRKFKNMGWOKLWVFTNQLFFYKSHODNHP LASPLPLGYSLTIPSESENIQKDYVFKLHPKSHVYFPAESEYTFERMEVIRATSS ASRPHVSHKESLVY"

## polyA\_site

BASE COUNT 864 a 952 c 927 g 699 t

## ORIGIN

Query Match 100.0%; Score 3442; DB 9; Length 3442;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 cccgcagccgcgcgcgtgtgagatattcttaagccgcttctcatcgagagaata 60  
 DB 1 CGCCGAGCCGCGCGCGTGTGGAGATATCTTAAGCGCGTTCATCATGGGAGAAATA 60

QY 61 gagcagagccgacccagagatcaagactcagctggggccgagaaattcgggagatcacc 120  
 DB 61 GAGCAGAGCCGACCCAGGATCAGACTGGGGCCCCCGGAAATTCGGGGATCAGTACC 120

QY 121 ttggaaactggaagagcgcgcgcacacacacacacacacacacacacacacacacac 180  
 DB 121 TTGGAACGTGGAGAGAGCGCGCCCAACACCTTCAGGAAACCTCGTGCCATCAAAATC 180

QY 181 cagatctgagatgacacccagagagcatttgaatttcaagaagactccttggaaggtg 240  
 DB 181 CAGATCTGGATGACACCCAGGAGGATTGTGAAGTTTCCAAAGAGCTCCTGGGAAGGTG 240

QY 241 ctgctgagatgagctttgcaacacacacacacacacacacacacacacacacacacac 300  
 DB 241 CTGCTGATGAGTTGTGCAACACCTCAACCTCGTGGAAAGGTGACTATTTTGGCCTCGAG 300

QY 301 ttctctgatcaaaaagatcacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 360  
 DB 301 TTCTCTGATCATAAAAGATCACGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGG 360

DB 301 TTCTCTGATCATAAAAGATCACGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGGCTGTGG 360  
 QY 361 attagaagccaaagcacgcttgttgaagttgttgtaaatcttcttcgcgctgaccac 420  
 DB 361 ATTAGAAGCCAAAGCACGCTTGTGTGAAGTTGTGGTGAATTTCTTCCGGCTGACCAC 420  
 QY 421 acacaactccaagaagaactcaaggttacctgttgcgctcagcagtgagcaggacttg 480  
 DB 421 ACACAACCTCCAAGAAGAACTCAAGGTACCTGTTGCGCTGCAGGTGAAGCAGGACTTG 480  
 QY 481 gctcaagcaggttgacgtgttaataacaccagcagcagctctcttgaatttcaacacattg 540  
 DB 481 GCTCAAGCAGGTTGACGTGTAAATGACACCGCAGCGAGCTCTCTTGATTTTACACATTTG 540  
 QY 541 caatctgagattgggattttgatgaagccttggacagagacacacttagcaaaaataaa 600  
 DB 541 CAATCTGAGATTGGGATTTTGTGAAGCCTTGGACAGAGACACTTACCAAAAAATAAA 600  
 QY 601 tacatacctcagcaagacgacactagaggacaaaatcgttgaatttcaaccataaccacatt 660  
 DB 601 TACATACCTCAGCAAGACGCACTAGAGGACAAAATCGTGGAAATTTCAACCATACCAATT 660  
 QY 661 gacaaaacaccagcagaatacagatttccagctccttagagattgcccgtcggctagagatg 720  
 DB 661 GGACAAAACACCAGCAAGATCAGATTTCCAGCTCCTAGAGATTGCCGCTCGGCTAGAGATG 720  
 QY 721 tatgaatccggttgcacccgcccagcagcagcagcagcagcagcagcagcagcagcagc 780  
 DB 721 TATGAATCCGGTTGCACCCGCCAAGGACAGGAGGACCAAGATCAATCTGCGCGT 780  
 QY 781 gccaacaggggaattctagtttccaggtttcacttaagatcaatgccttcaactgggccc 840  
 DB 781 GCCAACCGGGAATCTAGTGTTCAGGGTTTCACTAAGATCAATGTCCTTCAACTGGGCC 840  
 QY 841 aaggtgcggaagctgagcttcaagaggcgttcttctcaatcaagctccgcgcagagatgccc 900  
 DB 841 AAGGTGCGGAAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGTCCGGCAGATGCC 900  
 QY 901 aatagtcgtaccagatccttgaattccttgatggcagtcgggattcttgcagagtc 960  
 DB 901 AATAGTGGCTACAGATACCTTGAATTTCCGTGAGCGGCGGAGTTCTGCAAGTCC 960  
 QY 961 ttctgaaaaatcgtgtgaacatcatgccttctttagacttttgaagagcccaacca 1020  
 DB 961 TTCTGAAAAATCTGTGTGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCAACCA 1020  
 QY 1021 aagcccaagccgctcctttagcgggggtcatcatttcggttcagtggtgcgactcag 1080  
 DB 1021 AAGCCCAAGCCGCTCTCTTTAGCGGGGTTCATCATTTTCGGTTTCAGTGTGCGACTCAG 1080  
 QY 1081 aagcaggttctcagctatgttaagaagagagacataagaaggtgcagtttgaagaagag 1140  
 DB 1081 AAGCAGGTTCTCAGCTATGTTAAAGAGGAGGACATAGAAAGGTGCGCTTTGAAAGAG 1140  
 QY 1141 cacagaagattcattctatccgggagccttgcacagcctacagaactgaattcggaa 1200  
 DB 1141 CACAGCAAGATTCTATCTATCGGAGCCCTGCTTTCACAGCTTACAGAACTGAATTCGAA 1200  
 QY 1201 gtgctggagcagctcagcagagcaccagccttataatttggagaaggtgccgaatctca 1260  
 DB 1201 GTGCTGGAGCAGCTCTCAGCAGAGCACACGCTTACATTTGGAGAGGTCCGCAATCTCA 1260  
 QY 1261 gggggccagagctgcgcgcgagaaagaaacgaggtttccgcgaggagcgcggggtcg 1320  
 DB 1261 GGGGGCCAGAGTGTCCGCGCAGGAAAGAACCGAAGTTTCCGCCGGGAGCGGGGTG 1320  
 QY 1321 caccgagcccttgcgcgcgaggaagccccgcgggttaacaagcagcggagcggagccg 1380  
 DB 1321 CACCCGAGCCCTGCGCGCAGGAGAGAGCCCGGGGTAAACAAGCAGCGCAGCGAGCGCG 1380  
 QY 1381 tcgggccccacaggaagagagagaggtcggttaagatagaccacagagagtaaacct 1440  
 DB 1381 TCGGGCCCCACGGAGAGGAGGAGGTGCTGTTAAGGATAGGACCCAGCAGAGTAGAACCT 1440



QY 1441 cagcccccgagcccaagcacaggctccctgactgagcagtcctcaacttccagatgtct 1500  
Db 1441 CAGCCCCGCGAGCCCAAGCACAGGCTCCCTGACTGCGAGTCTCCTACCTTCCGAGTGCT 1500  
QY 1501 gtgaactcgaggggagtgagccctcccaactgagccttgctcccaactgagcccc 1560  
Db 1501 GTGAAGTCCGAGGGGAGTGAGCCCTCCCAACGTGACCTTGCTCTCCCAACCTGAGCCCC 1560  
QY 1561 gacacaaagcagggctctcccttgatcagccgctgctgaaagcagcctgccccgg 1620  
Db 1561 GACACAAAGCAGGCTCTCCCTTGATCAGCCGCTGCTGAATGACCAAGCCTGCCCCGG 1620  
QY 1621 acggacgatgaggtgagggccggaggagagattcccaactgataaagcgtactcata 1680  
Db 1621 ACGGACGATGAGGATGAGGGCCGGAGGAGAGATCCCAACTGATAAAGCGTACTTCATA 1680  
QY 1681 gctaaaggaagtgtctaccacagcagcgaacatatctgaagatctcgaagttatcactcg 1740  
Db 1681 GCTAAGGAAGTGTCTACCAAGCAGCGAATATCTGAAGGATCTCGAAGTTATCATTCTCG 1740  
QY 1741 tggtttcagagcacagttagcaaaagagcagccatgcccggaaagcactgaaaagtctcata 1800  
Db 1741 TGGTTTCAGAGCACAGTGAACAGAGGAGCCCATGCGGAGGACTGAAAAGTCTCATA 1800  
QY 1801 ttcccgaaatttgaaaccttgcaaaaaattcactataattttctcaaggaattgagcaa 1860  
Db 1801 TTCCCGAAATTTGAACCTTTGCACAAATTTTCATACTAATTTTCTCAAGGAATTTGAGCAA 1860  
QY 1861 cgaactgcccgtggaagcgcctcaaatcccaaaatcaagattaccagaagaatcgcc 1920  
Db 1861 CGACTTGCCCTGTGGAGGCGCGCTCAATGCCCAATCAGAGATTTACCAAGAAATCGCG 1920  
QY 1921 gatgctatgctgaagaacattcaggcatgaagcactggcgctcacctgtggaagcac 1980  
Db 1921 GATGTCATGCTGAACAACATTCAGGSCATGAAGCACTGGCGGCTCACCTGTGGAAGCAC 1980  
QY 1981 aqcgaggccttgaggccctggagaattggaatcaagagctcccgcgctggagaaacttc 2040  
Db 1981 AQCAGAGGCTTTGGAGGCGCTTGAGAAATGGAATCAAGAGCTCCCGCGGCTGGAGAACTTC 2040  
QY 2041 tcagagacttgagctacagaagtggttacctaccgctcaacacttccctcctggg 2100  
Db 2041 TCAGAGACTTTGAGCTGCAGAAGGTGTGTTACCTTACCCTCAACACCTTCTCCTCGGG 2100  
QY 2101 ccaactgcaacggctcatgcaactacaagcaggttcctggagcggtgtgcaaacaccacccg 2160  
Db 2101 CCACTTGCACCGGCTCATGCACCTACAAAGCAGTCTCTGGAGGCGCTGTGCAAAACACCCCG 2160  
QY 2161 ccgagccacgcgcacttcaggagctccgagccgcttggcagagatcacgagagtggtg 2220  
Db 2161 CCGAGCCACGCGGACTTCAGGGACTCGCGAGCGCTTTTGGCAGAGATCACGGAGATGGTG 2220  
QY 2221 gcacagctccagctacagatgaatgaatgagagatttccagaagctgcacgaactcaag 2280  
Db 2221 GCACAGCTCCACGGTACGATGATCAAGATGAGAGATTTCCAGAAGCTGCACGAATCAAG 2280  
QY 2281 aaagatttgatggcatgacaacttgggtccgggaagggaggttccatccgtctggcg 2340  
Db 2281 AAAGATTTGATTTGGCATTTGACAACTCTTGTTGTTCCGGGAAGGGAGTTATCCCTGGGC 2340  
QY 2341 agcctcagaactctcctggggaagggtccagcagcgcgatgttctcctgttcaacgac 2400  
Db 2341 AGCCTCAGCAAGCTCTCGGGGAAGGGGTCCAGCAGCGCATGTTCTTCTGTTCAACGAC 2400  
QY 2401 gtccctgtatcacagcgggggctgacgctcgaatcagtttaaaagtccacgggag 2460  
Db 2401 GTCCGTGTATACAGAGCGGGGGTGAAGGCTCCCAATCAGTTTAAAGTCCACGGGCGAG 2460  
QY 2461 ctcccgctctatggcatgacagatgagagagcgaagcagcagtgagggtgcccactgc 2520  
Db 2461 CTCGCCCTCTATGGCATGACGATTTGAGGAGACGGAGACAGTGGGGGTGCCCACTGC 2520

QY 2521 ctgaacctccggggccagcggcagtcctcatcatcgtggccgcccagttctcgtgccagatg 2580  
Db 2521 CTGACCTCCGGGGCCAGCGCAGTCCATCATCTGCTGGCCGCGAGTCTCTGGTCCGAGATG 2580  
QY 2581 gagaagtgggttgagacatccagattgaccattggcggaagaagcagcagcagcccc 2640  
Db 2581 GAGAAGTGGTTGAGGACATCCAGATGGCCATTGACCTGGCGGAGAGAGCAGAGCCCC 2640  
QY 2641 gcccttgagtctcctggcagcagcccccctgacaaagtccctctgataaagcccccgg 2700  
Db 2641 GCCCTTGAGTTCTTGGCCAGCAGCCCCCTGACAAAGTCCCTCTGATGAAGCCACCCGG 2700  
QY 2701 gctgaccaggaatcagaagatgaactgagcgcctcgcaacatcactgagcagcagccc 2760  
Db 2701 GCTGACCAGGAGTGAAGGATGACCTGAGCGCTCGGCACATCGCTGGAGCGCCAGGCC 2760  
QY 2761 ccgcaccgcggcaacacaaatggtgcagctgtgtgacccgcaacaccagcgtctccatg 2820  
Db 2761 CGGCACCGGGCAACACAAATGGTGCACGTGCTGGCACCAGCAACACAGCGTCTCCATG 2820  
QY 2821 gtggacttcagcatcgagtgaggaatcagttgtctggaacctgctgaggaaattcaaa 2880  
Db 2821 GTGGACTTCAGCATCGCAGTGGAGAAATCAGTTGTCTGGAACCTGCTGAGGAAATTCAAA 2880  
QY 2881 aacagcaacgggtggcagaagctgtgggtgtgtgttcacaaaacttctgctgttctctac 2940  
Db 2881 AACAGCAACGGGTGGCAGAACTGTGGGTGTGTTCACAAACTTCTGCGCTGTCTTCTAC 2940  
QY 2941 aaatcacaccaggaacaaatcctcccttgcagccttgcctcgtcgtcgtactcgtcaccc 3000  
Db 2941 AAATCACACCAGGACAAATCCTCCCTTGCACGCTGCCTCTGCTGGCTACTCTGCTCACC 3000  
QY 3001 atccctctgagtcggagaacatccagaagaactacgtgttcaagctgcactcgaagtc 3060  
Db 3001 ATCCCTCTGAGTCCGAGAAATCCAGAAAGACTACGTGTTCAAGCTGCACCTTCAAGTCC 3060  
QY 3061 caagtactacttcaggcgggaagcagctacagcttcgaagtggaatggaagtgc 3120  
Db 3061 CACGTCTACTTCTCAGGGCGGAAAGCGAGTACAGTTCGAAAGTGGATGGAAGTGC 3120  
QY 3121 ccagtgccaccagctctcctcgcgaccccgctgttagcacaagaagctctctgtg 3180  
Db 3121 CGAGTGCACCCAGCTCTGCTTCCGACCCCGACGCTGTGAGGCCACAAAGAGTCTCTGTG 3180  
QY 3181 tattgatggccggacacactcgtttccgcagctggttctcctggaagacgttctctt 3240  
Db 3181 TATTGATGGCCGGACACACTCGCTTTCCGCGAGTGGCTGCTTCTCTGGAAGACGTTCTCTTT 3240  
QY 3241 cttctgtattaatgaagcctgttaaaattaacacctgtctgaaaaatcaaaaacatggctt 3300  
Db 3241 CTTCTGTATTAAATGAAGCCTGTTAAATTTAAACCTGTCTGAAAAATCAAAAAATGGCTT 3300  
QY 3301 ccagcagctctcctgtctccacagcgcgcttttttaaccccgacctctcagcgtttgaa 3360  
Db 3301 CCCAGAGCTCTCCCTGTCTCCACAGCCCGGCTTTTAAACCCCGACCTCTCAGCGTTTGA 3360  
QY 3361 tgaacagcgtcccaactcagtcctcctggttcacgtcggggcgctgttctttagctagt 3420  
Db 3361 TGAACAGCGCTCCCACTCCAGTCTCTGCGATCCGCTGGGGCGCTGTTCTTCTTAGCTAGTG 3420  
QY 3421 ccagtattaaacattgtcatt 3442  
Db 3421 CCAGTATTAAACATTTGTCAAT 3442

## RESULT 2

BC004009 2632 bp mRNA ROD 12-JUL-2001  
LOCUS Mus musculus, clone IMAGE:3493093, mRNA, partial cds.  
DEFINITION  
ACCESSION BC004009  
VERSION BC004009.1 GI:13278387  
KEYWORDS house mouse.



```

Db      1261 TGACGGCTGGACACGCC 1278

RESULT  3
LOCUS   AB018336
DEFINITION Homo sapiens mRNA for KIAA0793 protein, complete cds.
ACCESSION AB018336
VERSION   AB018336.1 GI:3882306
KEYWORDS
SOURCE   Homo sapiens adult male brain cDNA to mRNA, clone_lib:pBluescriptII
SK plus clone:hk05692.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (sites)
AUTHORS  Nagase,T., Ishikawa,K., Suyama,M., Kikuno,R., Miyajima,N.,
Tanaka,A., Kotani,H., Nomura,N. and Ohara,O.
TITLE    Prediction of the coding sequences of unidentified human genes. XI.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro
JOURNAL  DNA Res. 5 (5), 277-286 (1998)
MEDLINE  99087487
REFERENCE 2 (bases 1 to 3997)
AUTHORS  Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.
TITLE    Direct Submission
JOURNAL  Submitted (08-OCT-1998) to the DDBJ/EMBL/GenBank databases. Osamu
Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology;
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
(E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913,
Fax:+81-438-52-3914)
FEATURES             Location/Qualifiers
     source            1..3997
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     clone             /clone="hk05692"
     clone_lib         /clone_lib="pBluescriptII SK plus"
     dev_stage         /dev_stage="adult"
     sex              /sex="male"
     tissue_type       /tissue_type="brain"
     misc_features     118..3282
                     /gene="KIAA0793"
     CDS               118..3282
                     /gene="KIAA0793"
                     /codon_start=1
                     /product="KIAA0793 protein"
                     /protein_id="BAA34513.1"
                     /db_xref="GI:3882307"
     translation       /translation="MGEIEGTYYVLQTAGMRLGAQTPVGVSTLEPGQTLPLPRMOEKLH
                     LRVKLLDNTMEFLPEKCDGVLLTVQMKRLNVCEDFYFGMEFQNTQSYWILPEP
                     KPIIRGRPKNVVLRAVLPDPQQLQCEYRVLFLQLRDLLEERLTCADTFA
                     ALLTSHLGQSEIGDQETDLDRHLKVNXYLPQGOHCLEKLEFPHOKHVGOTPAESDFQ
                     VLEARKLEWGYRFHMASDREGTKIQLAVSHMGVLVFOGTTKINTFNMSKVKPLSK
                     RKTFLKILPEVHGQDQDLEFLGSRDECKNFWKICVEVHTFTFLDQPKPAKAVFI
                     FSKSGSFRYSGRYTKQVLDFKDGMMKRIYERRHSTKTSVSRALTAADLPKQISIPPE
                     GLRTPASANAFYSLSPSTLVSPGLPEFKDSSSLTDQVSYVKSPPAAERSGVA
                     GDDTPTSAQGLPDPALQPGVLTSTKSPQSPSRKSPLSLSPAFQVPLGPAEQGSPFL
                     LSPVLDAGGACMGCEPRHKRVPDAEAVFIKVEILATERTYLDLEVITVWFRSAVV
                     KEDAMPTALMTLLFSNIDTYEPHGRFLREVEQRLALWELGSKAHKPGSHQRIQDILL
                     RNRQLKFTSYFORHDEVULTELEKATCKKCKLEAVYKFELOQVCIPLNTPFLKPI
                     QRLLHYLNTLRGCEYSGPHGDYADCHDAKAITVTTTQHLILRLENLQKLETLQ
                     RLQVLENTJAPGREYIRRECKHLKTKGLQRMFFLFSDMLLYTSGVAGTSGHFRI
                     GLLPALQVLEEDSNWSPHCFYIVAAQKTIIVAASTRLEKEKMWLDLNSAIOAKS
                     GGDTALPCRTVCTPPRPSNPVSLQEEDDAGVRSSLEGHGHRANTTWHCVQ
                     RNTVSGRADHSAVENQLSGYLLRKFNKSHGWOKLWVYTFNFCLEFPYKTHQDDYPPLAS
                     LPLHGTSIPRSGDGHKDYKFEKLFQKSHVYFRAESKYTFERNMEVIQGASSSAG
                     APSIVDDGPQSPGLGEMVRGKE"
     998 a 1132 c 1068 g 799 t
BASE COUNT
ORIGIN

```

Query Match 30.18; Score 1037.4; DB 9; Length 3997;

	Best Local Similarity	60.28;	Pred. No. 6.4e-227;	Mismatches 1864;	Conservative	0;	Mismatches 1186;	Indels	45;	Gaps	7;
Qy	69	gccgaccacagatcacgaactcgggggccccggaaaaattcggggatacagtacccttggaaacg	128								
Db	150	GCACACTGCCAGGATGGCTTGGGTGGCCCGCAGACCCTCTGGGAGTTAGCACCCCTTGAGCC	209								
Qy	129	tggacagaagccgcgcccaaaccttcaggaaaactcgtgtccatacaaaatccagatgct	188								
Db	210	TGGCAGACTCTCTTGCCCAAAATGCAAGAAGAACCCATGCACTTCAGAGTAAGCTGCT	269								
Qy	189	gqatgacaccacagagcatttgaaattccacaagagctcctgggaaggctgactgga	248								
Db	270	GGACAACACCATGGAATAATTGTGACATTGAGCCTAAATTCGATGGCCAGGTATTACTGAC	329								
Qy	249	tgcagtttgaacacacctcaacctcgttgaaggtagactattttggcctcgaattttcctga	308								
Db	330	ACAAGTGTGGRAGCGTTAAACCTGGTAGAATGTGACTACTTCGGGATGGAGTTTCANAA	389								
Qy	309	tcacaaaagatcacggttgtgtgtgga tctcctaiaaaccacttg tgaacacagattagaag	368								
Db	390	TACTCAGTCCTACTGGATTGTGGCTTGAACCTATGAAAACCCATCATTAGGCAATACGAG	449								
Qy	369	gcaaagcacgdtgtttaaagtttgttgtgaaatttcttcogctgaccacacaact	428								
Db	450	GCCAAGAAATGTGTGCTTCGCTTAGCTAGCTGTAAAAATTTTTTCACCTGATCTCTGGTCAGT	509								
Qy	429	ccaagaaacctcacaaagtacctgtcgtcgtcaggtgaaagcaggacttggtccaag	488								
Db	510	ACAAGAAGATATACAAGATACTGTTTGGCTTGGCACTTAAGAGAGACCTGCTGGAAGA	569								
Qy	489	caggttgacgttgtaatgacacacgcgcagctctcttgatttcaacatatgtgcaatctga	548								
Db	570	CGCTTTGACCTGTGTGACACACACAGCGGCCCTTCTCACGTCCCCTCTCTGACGTGCGA	629								
Qy	549	gatlggggattttgatgaagccttggacagagagcaacttagcaaaaaataaatatcatacc	608								
Db	630	AATAGGAGATTACGATGAACCGCTGGACCGAGAGCACCTCAAAAGTGAACGAGTATTGGC	689								
Qy	609	tcagcaagacgcactagaggaacaaactcgttgaatttcaccataaccacattggacaaac	668								
Db	690	TGGCCAGCAGCACATGCCCTTGAGAAGATCTAGAAATTCCTACAGAAGCAGCTGGGCCAGAC	749								
Qy	669	accagcagaatcacgattccagctcctagagattgcccgtcggttagagatgtatggaat	728								
Db	750	ACCTGCTCAGTCGGATTTCCAGGTGCTCGAAATTCGCTCGAAAGTTGGAAATGCTACGGCAT	809								
Qy	729	ccggttgaccccgccaaaggacagaggaaggcaagatcaactcgtggccggttgcacaacac	788								
Db	810	CAGATTTTCCATGGCTTCTGACAGGAAAGAACCAAGATTCACCTGGCAGTTTCCACAT	869								
Qy	789	gggaattctagrtttcagggtttcactaaagataaatcccttcaactcaggcccaagtgcg	848								
Db	870	GGGTGACTCGTGTTCAGGCGACCCACCAATCAACACTTTTCACTGGTCCCAAGTCGG	929								
Qy	849	gaagctgagcttcaagaggaagcgctttctcatcaagctccggcgacagatgcacaatagtgc	908								
Db	930	TAACTAAGCTTCAAGAGGAAAAGATTCTTATCAAACCTTCATCCAGAGTTTCATGGACC	989								
Qy	909	gtaccagataaccttggaaattcctgatggccagtcgggatttctgtcaagtctcttgaa	968								
Db	990	TTACCAGGACACATTAGAAATTTTGTGGGTAGTAGAGATGAATGTAAGAACCTTCTGGAA	1049								
Qy	969	aatctgtgttgaacatacatgctctctttagactttttgaagagcccaacaaagcccaaa	1028								
Db	1050	GATTTGTGTGGAGATACACACTTTTTTGTAGACTTTTGTGACCAACCTTAAGCCCCAAAACAAA	1109								
Qy	1029	gccgctctctttagccgggggtcatactttcgggttcagtgctcgagctcagaagcaggt	1088								
Db	1110	AGCGCTTCTTTCAGCCGGGGCTCCTCCTTCAGATACAGTGGAAAGAACTCAGAAACAAC	1169								
Qy	1089	tctcgactgtttaaagaaggaggacataaagaaggtgcagtttgaagaagcacaagcaaa	1148								

Db 1170 AGTAGATTATTTCAAAGACAGTGAATGAAGAAATTCATATCAAAAGAGGCACAGCAA 1229  
Qy 1149 gattcatttatccggagccttgcttcacagcctacagaaactaattcggaaagtgtgga 1208  
Db 1230 GACCCACAGTCC-----GTTGAGGTCTGACTGACAGACTACCAAA 1271  
Qy 1209 gcagttccagcagcagcagccttacatttgagaaagtgccgaatctccagggggccca 1268  
Db 1272 ACAGAGCATCTCATTCCTCCGAGGATGAGGACTCTCCCTCCCATCTTCACGCAATGC 1331  
Qy 1269 gaactgcccggcgggaaagaaacgaaggtttccgcggggagccggggtcgaccgccag 1328  
Db 1332 CTTTATCTCGCTCTCTCCCTCCACTCTGGTCCCTCTGGCTCCGAGTTTAAAGACAG 1391  
Qy 1329 cctcgccgagagagagcccgccgggttaacaaagcggcggagcagccctcgccgccc 1388  
Db 1392 CAGCAGCTCCCTCACAGATCCCAAGTTCCTTACGTACAGAGTCCAGCTGCAGAGAGGG 1451  
Qy 1389 cagggaggagagagagag--tcgttaaggatagagccagcagagtaaaacctcagcc 1445  
Db 1452 CAGTGGAGCAGTGGCTGGAGGCCCGACACACCATCGGCCCGCCCTCGGGGCCCCCGC 1511  
Qy 1446 ccgcgagcaagcaagggctccctgactggcagtcctcactcacttcgcagctgtctgtaa 1505  
Db 1512 ACTCCAGCCTGGTCCAGAGCCTTTCCACGAAGAGTCCCTCAGCCTTCTCCCTCCAGCGGAA 1571  
Qy 1506 ctgcgagggggagtgccctccaaagtgacctgtctccaaacctgaagcccgagcac 1565  
Db 1572 GAGCCCTGTAGTCTGAGCCCTGCATTTTCAGGTGCGCTT-----TGGGCCCACTGA 1622  
Qy 1566 caagcaggcctctcccttgatcagccgcgtgctgtaagaccagcgctgccccggcagga 1625  
Db 1623 ACAGGCTCATCCCACTCTCTGAGCCCTGTCTCTAGTGATGTGGCGGAGCCGGGATGA 1682  
Qy 1626 cgaatgagatgagggccgaggaagagattcccaactgataaagcgaactcacttaactaa 1685  
Db 1683 CTGCGAG--GAGCCAGACACAAAGCCGCTGCTGCGAGACAGAGCCCTACTTATATGCAA 1739  
Qy 1686 ggaagtgtctaccacccagcagacatatctgaaggtatctcgaagttatcactctggtgt 1745  
Db 1740 AGAGATTCTCGTACAGACGAACATACCTCAAGGATTTAGAGTTATTTACCGTGTGGTT 1799  
Qy 1746 tcagagcacagttagcaagagagaccatgccgggaagcactgaaagtctcattatcc 1805  
Db 1800 CCGCAGCGCAGTGGTGAAGGAGGAGCCCATGCTGCGAGCTCTGATGACGCTGCTCTCTC 1859  
Qy 1806 gaatttgaaactttgcacaaatttctaactaattttctcaaggaaattgagcaagcact 1865  
Db 1860 CAACATCGATFCCCATCTATGAGTTCACAGAGGCTTCTTGGCGGAGGTGGAGAGGCT 1919  
Qy 1866 tgccttggtggaaggccgctcaaatgcccaaatcagag--attaccaagaatcgcca 1922  
Db 1920 GGCACCTCTGGGAGGCGCCCTCCAAAGCCACACAAAGGACGATCATCAAGATCGGGA 1979  
Qy 1923 tgcattctgaagaacattcagggcattgaagcactggcggtcactcctggaagcacag 1982  
Db 1980 CATCTCTCTCAGGAACATCGCCAGTAAAGGAGTTTACAGCTACTTCCAAAGACATCA 2039  
Qy 1983 cgaagccttgagccctgagagaatgaatcaagaactcccgcggtgagaaactctg 2042  
Db 2040 CGAGGCTTCAACAGAACTGGAAAGGCTACCAAGCCTGTAAAGATTTGGAGGACGTGA 2099  
Qy 2043 cagagactttgagtcgagaaggtgtgttacctaccgctcaacaccttctcctcgccc 2102  
Db 2100 CAAGGATTTGAGTTCGAGAGGTCTGCTACTTGTCTCTCAACAGCTTCTCTGAGGCC 2159  
Qy 2103 actgcaccggtctatgcatacaagcaggtccttgagcggcgtgtgcaaacaccaccgcc 2162  
Db 2160 CATCCAGCGCTGCTGCATACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2219  
Qy 2163 gaggcagccgacttcagagactccgagccgctttgagagatacagcagagatggtgc 2222  
Db 2220 CGGGCACCATGACTAGCTGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2279

Qy 2223 acagctccacgggtacgatgatcaagaatgtcagaagctcgaagaactcaagaa 2282  
Db 2280 CACACTACAGCACAATCTCTATCCGGCTGGAGAACCTGCAGAACCTACGGAGGTGACGC 2339  
Qy 2283 agatttgattggcattgacaaatcttggttccgggaaaggagttcactcgtctggcag 2342  
Db 2340 GGACCTGGTGGGCATAGAGAACCTCATTTGCTCTCGCAGGGAGTTTCATCGTGGGCTG 2399  
Qy 2343 cctcagaagcctctcggggaagggttcacagcagcagcatgttctcctgttcaacagct 2402  
Db 2400 CCTTCAAGAGCTACCAAGAGGGCCTGCAGCAGAGGATGTTTCTGTCTTCAGATAT 2459  
Qy 2403 cctgtatatacagcagccggggcctgagccctcactcagtttaaggtccacagggcag 2462  
Db 2460 GTTGTCTGTACACAAAGAGGAGTTGCAGGGACCAACCCTCCGGATCCGGGGCTCCT 2519  
Qy 2463 ccgctctatggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2522  
Db 2520 TCCCTTCCAGGCGATCTGTGTGAAGAAAGTGATAACGAGTGTCTGTTCACACATGTTT 2579  
Qy 2523 gaccttccgggcccagcggcagtcactcactcactcactcactcactcactcactcact 2582  
Db 2580 CACCATCTACGGCGCTCAGAAACAAATCGTGTGGCAGCCAGCACCTCGGCTGGAGAAAG 2639  
Qy 2583 gaagtgggttgagggacatccagatggccttgccttgccttgccttgccttgccttgcct 2642  
Db 2640 GAAGTGGATGCTGGACCTGAATCCCGGATCCNAGCAGCCAGAGTGGCGGTACACGGC 2699  
Qy 2643 ccttgagttcctggccagca-----gccccctgacaacaagctcccctgatgaagccac 2696  
Db 2700 CCTGTCACTGCCAGGCGGCACCTGTGTGCACTGTCTCCCCCAGATCCCCAACAGGTATC 2759  
Qy 2697 cgcgctgaccaggagtcagagagatgacctgagcgcctcgcgcacatcgcctgagcgcaa 2756  
Db 2760 TCTG---GAGCAGGAGTCAAGATGATCTCGGGGTGTCCGCGAGCTCCCTTGGAGGGCA 2816  
Qy 2757 ggc 2816  
Db 2817 TGGCCAGCAGCCGGGCCAACACCAATGACGCTGTCTGTGTGTGTGTGTGTGTGTGTGTGT 2876  
Qy 2817 catgttgacttgaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2876  
Db 2877 CAGGCGAGACCACTGACGCTGTCGAGAACCAAGCTTTTTCAGGATATCTGTAAGAAGTT 2936  
Qy 2877 caaaacacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2936  
Db 2937 CAAAACAGCTATGCTGGCGAGAGCTCTGGGTGCTTGTACCAACTTCTGTTGTCTT 2996  
Qy 2937 ctacaaatcacaccaggaacaatcaatcccttgcagcctgcctcctgcctgcctgcctgcct 2996  
Db 2997 CTACAAAACATCATCAGGATGACTACCCACTGGCCAGCCCTCCCGCTGCTGGCTACAGCGT 3056  
Qy 2997 caccatccctctgagtcgagagaacatccagaagaactcagtggttcaagctgcacttcaa 3056  
Db 3057 GAGCATCCCCAGGAGGCGGATGGCATACACAAGACTATGTTTTCAGCTCCAGTCAA 3116  
Qy 3057 gtccacgctcactacttcagcggcgaagcagcagcagcagcagcagcagcagcagcagcagc 3116  
Db 3117 ATCCCAAGCTACTTCTTCCGGCTGAGAGCAAGTACACATTTGAAGAGGTGGATGAGGT 3176  
Qy 3117 gatccgagtcgacccagcctgcctcgcgacccc 3151  
Db 3177 GATCCAGGGGCGCAGCAGCTACGCCGGAGGGGCC 3211

## RESULT 4

BC009153

LOCUS

DEFINITION

ACCESSION

VERSION

BC009153 3719 bp mRNA ROD 12-JUL-2001  
Mus musculus, Similar to KIAA0793 gene product, clone MGC:6304  
IMAGE:2655209, mRNA, complete cds.  
BC009153  
BC009153.1 GI:14318718

KEYWORDS	MGC.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	1 (bases 1 to 3719)
JOURNAL	Direct Submission
REMARK	Submitted (05-JUN-2001) National Institutes of Health, Mammalian
COMMENT	Gene Collection (MGC), Cancer Genomics Office, National Cancer
	Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
	USA
	NIH-MGC Project URL: <a href="http://mgc.nci.nih.gov">http://mgc.nci.nih.gov</a>
	Contact: MGC help desk
	Email: <a href="mailto:cgapps-f@mail.nih.gov">cgapps-f@mail.nih.gov</a>
	Tissue Procurement: Gilbert Smith, Ph.D.
	cDNA Library Preparation: Life Technologies, Inc.
	cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
	DNA Sequencing by: Baylor College of Medicine Human Genome
	Sequencing Center
	Center code: BCM-HGSC
	Web site: <a href="http://www.hgsc.bcm.tmc.edu/cdna/">http://www.hgsc.bcm.tmc.edu/cdna/</a>
	Contact: <a href="mailto:villalon@bcm.tmc.edu">villalon@bcm.tmc.edu</a>
	Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,
	A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
	Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK Plate: 5 Row: m Column: 20.

```

FEATURES
source
i...3719
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="MGC:6304 IMAGE:2655209"
/tissue_type="Mammary tumor, MMTV-LTR/INT3 model, 5 month
old mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Mam2"
/lab_host="DH10B"
/notes="Vector: pCMV-SPORT6"
99...3296
CDS

```

```

99..:3296
/codon_start=1
/product="Similar to KIAA0793 gene product"
/protein_id="AAH09153.1"
/db_xref="GI:14318719"
/translation="MGEISGTVALPTSGTRGGOTAIGVSTLEPEQSLSPRMOEKHM
RIVRLKLDSTVELDIEPKDCGOVLLTVQVKHNLIECDYFGLERFNQSVYWLHPQM
KPIIRVRAPKNALVAKVDFPDQEGQELRYFLVALQRLDLLEERLTLTAMTA
ALLISHLDEIGDGYDETLDREHKANLYPNQEKSLKILDFHQRHTQTPAESDFQ
VLEIARKLMEYGIHFIMASDRECTKINLAVSHMGVLVFGQTQKINTFNKSVKRLSFK
RKRFLLFVHGVPQDYLLEFLGLSRDECKFEWKICVEYHTEFLSDOPKKQKAVF
FGSGSFYSGRTKQGLVYVDKGMKRIPIYERRHKTRTSLHALTVDLPLKQSVSTF
GURTSASLANSANVSFYPPSSSLSPGLNLKDSSSLVDPOQAPVIKSTAERSSGGS
SDGPTQSANHLGPVPLRQPGFGFSDMSQPSPPSSLSKLSLCPQLQAALSTBAQGS
PVLPSVLSGALTPMDQBEQRKHKMPDEAYFAIKELATERTYIKLDEVTYIWFERS
VLIKESAMPAALMALFISLNIIDPYVEFHRGFLHEVEQRLALWEGFSAHKLGDHRIOT
ILLRNRRQKLEFTSYQRHDEVILETEKATKHCKLEAVYKEFELQACVCLPNTLTL
KQVORLVHYRLILSLCAHYSGPHRDYADCHEALKAITVTELOQSLTLENLOKLT
ELQDRLGVENLTAPGRTRETCGLHKLTKGKLOQRMFLFSDMLLYTSKSVTGASHF
RIGCFPLRGMLVEESENWSYLHCFIYAAOKTIYVAASTRELEKEMQDNLNAAQA
AKTIGDSPVLVGLGGPYITRPSDEVSLEESDGRGNRLEGNSGHRANTHWCVC
YNTISYVRADHASAENVQTLISLLRKFKNKNGWKLWVFTNEFLFYKTHQDDYPLA
SUPPLGYSYSLPREADSIHKDYVFKLQFHSVYFNFLEAESKYTERWMDVITKRASSPG
RPPSFYQDCSHHSQPEATIREACEAPCPLDKNL"
997 a 979 c 929 g 814 t

```

Query Match	29.68;	Score 1019.2;	DB 10;	Length 3719;
Best Local Similarity	59.1%;	Pred. No. 9.4e-223;		
Matches 1867;	Conservative	0;	Mismatches 1263;	Indels 30;
				Gaps 6;

Qy 70 ccgacccaggatcacgactggggggcccgaaaattcggggatcagttaccttggaacgt 129

Db	132	CAAC	TT	CAGG	GAC	ACG	CT	TGG	TGG	TGCC	AG	ACCG	CA	CT	TGG	AG	TAG	TAC	CT	GG	CC	A	191
QY	130	ggac	agac	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	cc	189
Db	192	GAC	CAG	AG	CT	GT	CAC	CC	AG	AG	TGC	AG	GAG	AG	CAC	AT	CG	CAT	CAG	AG	TGA	GC	251
QY	190	gat	gac	cacc	ag	gag	cat	t	ta	g	at	t	cc	a	a	a	a	g	at	c	ct	cc	249
Db	252	GAC	AG	CA	CT	G	TAG	AT	T	TT	G	ACA	T	CG	AG	CG	GA	AG	T	G	T	AG	311
QY	250	g	ag	tt	g	ca	cc	ac	ct	ca	ac	ct	cg	tg	ga	ag	tg	a	ct	at	tt	g	309
Db	312	CA	AG	T	G	GA	AG	CA	T	T	GA	A	T	TC	A	T	GA	A	T	TC	A	T	371
QY	310	c	a	a	a	a	a	g	a	t	c	ac	cc	at	g	t	g	a	a	c	a	g	369
Db	372	GT	CC	AG	T	CT	ACT	AG	T	T	GG	CT	T	GA	AC	CT	A	T	GA	AC	CA	T	431
QY	370	c	a	a	a	c	g	t	t	g	t	t	a	g	t	t	c	g	t	t	c	g	429
Db	432	CC	A	A	A	A	A	T	CG	GT	CT	CG	CA	G	T	A	A	A	T	T	T	T	491
QY	430	ca	a	g	a	a	a	c	t	c	a	a	g	t	g	c	g	t	g	c	a	a	489
Db	492	CA	A	A	A	A	G	A	G	T	A	C	C	T	G	T	T	G	C	T	T	G	551
QY	490	a	g	t	g	a	c	t	g	a	a	c	g	a	c	a	c	t	c	t	g	a	549
Db	552	CG	CT	G	A	CT	G	C	AG	CC	A	AC	T	G	AG	CC	CT	T	C	A	T	611	
QY	550	a	t	t	g	g	g	a	t	t	g	a	a	c	t	t	g	a	a	c	t	g	609
Db	612	AT	CG	G	A	G	A	T	T	A	T	G	A	A	C	C	T	C	A	A	A	671	
QY	610	c	a	g	a	a	c	a	c	t	a	g	a	a	a	a	a	a	a	a	a	a	669
Db	672	A	A	C	C	A	G	A	A	A	T	C	T	A	G	A	A	A	G	A	T	731	
QY	670	cc	a	g	a	a	a	t	c	a	g	a	t	t	c	a	g	a	t	t	c	a	729
Db	732	C	T	G	C	A	G	A	G	T	T	C	C	A	G	T	T	G	A	T	T	G	791
QY	730	c	g	t	t	c	a	c	c	g	c	a	a	g	a	a	c	a	a	a	a	a	789
Db	792	A	G	T	T	C	A	T	G	C	T	C	A	C	A	G	A	G	G	C	A	A	851
QY	790	g	a	a	t	c	t	a	g	t	t	c	a	a	a	a	a	a	a	a	a	a	849
Db	852	G	T	G	T	C	T	G	T	T	C	C	A	G	G	T	A	C	A	C	A	A	911
QY	850	a	a	c	t	a	g	c	t	c	a	a	g	a	a	g	c	t	t	c	a	a	909
Db	912	A	A	A	C	T	T	C	A	A	A	A	A	A	A	A	A	A	A	A	A	A	971
QY	910	t	a	c	c	a	g	a	t	c	c	t	g	a	t	t	c	g	a	t	t	c	969
Db	972	T	A	C	C	A	G	A	C	A	G	A	G	A	T	G	A	G	A	C	T	T	1031
QY	970	a	t	c	t	g	t	t	a	a	c	t	c	t	t	t							





SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
1 (bases 1 to 2888) Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Bloeker, H., Boecher, M., Brandt, P., Mewes, H.W., Weil, B. and Wiemann, S.  
TITLE Direct Submission  
JOURNAL Submitted (15-MAR-2000) MIPS, Am Klopferspitze 18a, D-82152 Martinsried, GERMANY  
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by GBF (National Research Centre for Biotechnology Ltd., Braunschweig/Germany) within the cDNA sequencing consortium of the German Genome Project.  
This clone (DKFZp762P046) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de. Further information about the clone and the sequencing project is available at <http://www.mips.biochem.mpg.de/proj/cDNA/>.  
FEATURES  
Location/Qualifiers  
1..2888  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/clone="DKFZp762P046"  
/clone\_lib="762 (synonym: hmcl2). Vector pSport1; host DH10B; sites NotI + SalI"  
/dev\_stage="adult"  
/tissue\_type="melanoma (MeWo cell line)"  
polyA\_signal 2852..2857  
polyA\_site 2859  
BASE COUNT 879 a 592 c 620 g 797 t  
ORIGIN  
Query Match 10.0%; Score 342.8; DB 9; Length 2888;  
Best Local Similarity 68.2%; Pred. No. 7.5e-68;  
Matches 476; Conservative 0; Mismatches 222; Indels 0; Gaps 0;  
QY 445 aggtacctgttcgctgcagtgagcaggactgtgctcaaggcaggtgacgtgtaaat 504  
DB 123 AGGTATCTTTTACTCTCAATAAAGAGGATTTGGCTCTAGGAGGCTTCCATGCAGT 182  
QY 505 gacacagcgcagctcttatttcacacattgtgcaatctgagatctgggtattttgat 564  
DB 183 GACAACTGTACAGCGTGTGATGTCACATCTTACAACTCAGAACTTGGAGACTTTTAT 242  
QY 565 gaagccttgacagagacacttagcaaaaaataataacatacctcagcaagcagcacta 624  
DB 243 GAAGAAACAGATAGGAAGCATCTGGCACAACCTCGGTACTTACCAACCAAGACTGTTTA 302  
QY 625 gaggcaaaatcgtggaatttcacataaccacattggacaaacacacagcagaatcagat 684  
DB 303 GAGGCAAGATCATGCATCTTATCAGAGCAGATTTGGCAGGAGCCACTGAATCTGAC 362  
QY 685 ttccagctcctagagattgcccgtggctagagatgtatggaatccggttgaccggccc 744  
DB 363 ATTCTGCTACTGGACATAGCAAGGAGCTGGATATGTATGGCATCAGGCTCACCCCGCC 422  
QY 745 aaggacaggaagcagagatcaatctggcgttgccaaacacaggaattctagtgttt 804  
DB 423 AGTAGGTTGAAGGGATGAGCATTCACCTGGCTGTGCTCACATGGGAGTATGGTGTTA 482  
QY 805 caggggttcactaaagatcaatgcttcaactggccaaagtgccgaagctgagcttcaag 864  
DB 483 CGGGGAATACAAAGATCAATCTTTACTGGGCTAAATCCGCAAGTGTAGTTTAAAG 542  
QY 865 aggaagcgtttctctcaagctccggccagatgccaatagtgcgtaccagataaccttg 924  
DB 543 AGAAAGCATTTCTCATCAAACTTCATGCCAATATCTTGTTGTTGGCAAGGATACCTTG 602  
QY 925 gaattcctgatggcagtcgggatttcttgcgaagtccttcttggaataatctgtgtgaacat 984

Db 603 GAGTTCACCATGCCAGCCGAGATGCTCAAGGCTTTCTGGAAGACTTGTGTGGAAATAC 562  
QY 985 catgcctcttttagacttttttaagagagcccaaaccaagcccaagccgtctctttagc 1044  
Db 663 CATGCTTCTTCAGGCTTTCCGGAAGAGAGCCCAATCAAGCCCAAAACCTACTCTCAGC 722  
QY 1045 cgggggtcatctatttcggttcagtgctcggactcagaagcaggttctcactatgttaa 1104  
Db 723 AAGGTTTCAGGTTTTCGGCTATAGTGGAGCAACCCAAAGGCACTTTTGAATATGGAGA 782  
QY 1105 gaagagagacataagaagtgctcagtttgaagaagca 1142  
Db 783 AAAGGAGGCTGAAGAGCTTGCCATTTGAAGGAACA 820  
RESULT 6  
AL137249  
LOCUS AL137249 181800 bp DNA HTG 03-AUG-2001  
DEFINITION Homo sapiens chromosome 13 clone RP11-111L24 map q31.3-32.3, \*\*\*  
SEQUENCING IN PROGRESS \*\*\*, 4 unordered pieces.  
ACCESSION AL137249  
VERSION AL137249.19 GI:15131444  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
1 (bases 1 to 181800) Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Garner, P.  
TITLE Direct Submission  
JOURNAL Submitted (02-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
requests: [clonerequest@sanger.ac.uk](mailto:clonerequest@sanger.ac.uk)  
On Aug 9, 2001 this sequence version replaced gi:15131193.  
COMMENT  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: <http://www.sanger.ac.uk>  
Contact: [humquery@sanger.ac.uk](mailto:humquery@sanger.ac.uk)  
----- Project Information  
Center project name: BA31K22  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 180516 bases at least Q40  
Consensus quality: 180899 bases at least Q30  
Consensus quality: 181168 bases at least Q20  
Insert size: 181500; sum-of-ctigs  
Insert size: 174530; 1.8% error; agarose-fp  
Quality coverage: 6.26x in Q20 bases; sum-of-ctigs Quality  
coverage: 6.72x in Q20 bases; agarose-fp  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 4 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 114302: contig of 114302 bp in length  
\* 114303 114402: gap of 100 bp  
\* 114403 135098: contig of 20696 bp in length  
\* 135099 135198: gap of 100 bp  
\* 135199 173346: contig of 38148 bp in length  
\* 173347 173446: gap of 100 bp  
\* 173447 181800: contig of 8354 bp in length.  
\* Location/Qualifiers  
1..181800  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
FEATURES  
source

Whitehead Institute for Biomedical Research  
9 Cambridge Center, Cambridge MA 02142 USA  
Tel: 617 252 1900  
Fax: 617 252 1902

Primer A: ATGTTTGTGATTTTCAGACAGGTG  
Primer B: GAGTCTCTTGTTGATGATGCC  
STS size: 127

Presoak:  
Denaturation: 56 d  
Annealing: 56 d  
Polymerization:  
PCR Cycles: 35  
Thermal Cycler:  
Protocol:  
Template: 10 ng  
Primer: each 5 pM  
dNTPs: each 4 nM  
Taq Polymerase: 0  
Total Vol: 20 ul

Buffer:  
MgCl<sub>2</sub>: 1.5 mM  
KCl: 50 mM  
Tris-HCl: 10 mM  
pH: 9.3

Derived from dbEST (genbank accession T81406).

```

FEATURES                                Location/Qualifiers
     source                             1..483
                                     /organism="Homo sapiens"
                                     /db_xref="taxon:9606"
                                     /map="291.8 cR from top of Chr13 linkage group"
     SPS                                17..143
     primer_bind                        17..40
     primer_bind                        complement(121..143)
     BASE COUNT                        121 a 103 c 122 g 127 t 10 others
     ORIGIN

```

Query Match 9.4%; Score 322.2; DB 11; Length 483;  
Best Local Similarity 90.1%; Pred. NO. 4.4e-63;  
Matches 365; Conservative 0; Mismatches 34; Indels 6; Gaps 2;

QY	2904	gtgggtgtgtttcaacaaactctgcctgttcttctacaaat-----cacaccagagacaatc	2959
Db	414	GTGGGTGTGTTCANAAAAATTTCTGCCTGGTGCTTCTTANAATTTNAACACCCAGGNACAATC	355
QY	2960	atccccttgcca--gcctgcctctctgcctgcgtactcgcctcaaccatccctctgagtccga	3017
Db	354	ATCCCCCTTTGCAAGCCCTGCCTTCTGCTCGGNTACTCGCTACCCATCCCTCTGAGTCCGA	295
QY	3018	gaacatccagaagaactacgtgttcaagctgcacttcaagtccacagctctactacttcag	3077
Db	294	GAACATCCAGAAGAAGACTACGTGTTCAAGCTGCATCTCAAGTCCCAACGCTACTACTTCAG	235
QY	3078	ggcggaaagcaggtacacgcttcgaaggttgatggaagtgcacgcagtgccaccagctc	3137
Db	234	GGCGGAAAGCGAGTACACGTTCCGAAAGTGGATGGAAGTGATCCGCACTGCCACCAAGCTC	175
QY	3138	tgcctcgcagccccacgctgttagcaccacaagaagtctcttgtgtatgtaggcgcggacac	3197
Db	174	TGCTCGCACCCACCGTGTGAGCCACAAGAGAGTCTTCTGTGATTGATGGCGCGGANAC	115
QY	3198	actcgtttccgcagtggtcgttcttcctggaagacgttcccttctctctgtatttaataagag	3257
Db	114	ACTCTGTTCCGCAGTGGCGTCTTTCTCTGGAGACAGTTTCCCTTCTCTGTATTAAATGAAC	55
QY	3258	ctcgtgaaataatacacctgtctgaaatacaaaacatggcttcc	3302
Db	54	CTGGGTAAATTTAAACACCTGTCTGAAATCAAAAACATGNTTCCC	10

misc_feature	1. 114302	/note="assembly_fragment:00010 fragment_chain:1 clone_end:SP6 vector_side:left" 114403. 135098 /note="assembly_fragment:01859 fragment_chain:1" 135199. 173346 /note="assembly_fragment:02403 fragment_chain:1" 173447. 181800 /note="assembly_fragment:02166 fragment_chain:1 clone_end:T7 vector_side:right"	301 others
misc_feature			
misc_feature			
misc_feature			
BASE COUNT	49780 a	43544 c	41179 g
ORIGIN	46996 t		

Query Match	9.9%	Score 339.2	DB 2	Length 181800
Best Local Similarity	96.4%	Pred. No. 3.8e-67		
Matches 347	Conservative 0	Mismatches 13	Indels 0	Gaps 0
QY 3083	aaagcaggtcacgattcgaaagggtggatgaagatgattccgcagtcgagtcgccaccagctctgcct	3142		
DB 89227				
QY 3143	cgcgaccaccagtttdagccacaagaagtcctcttgattgattggtgcgggacacactcg	3202		
DB 89287	CGCGACCCACGCTTTCAGTCAACAAGAGTCTCTTGTGATTGATGGCCGGACACACTCG	89346		
QY 3203	tttcgcgagtgtgctgttctcgtggaagacgcttctcttctctgtattaatgaagcctgg	3262		
DB 89347	TTTCCGCAGTGGCTGCTTTCTCTGGAAGACGTTTCTCTTCTCTATTATGAAGCCTGG	89406		
QY 3263	taaaattaacacctgtctgaaaaatcaaaaacatggtctccacgacgctctcctgtctcca	3322		
DB 89407	TAAAAATTAAACACCTGTCTGAAATCAAAAAATGGCTTCCAGCAGAGTCTCTCTGTCTCCA	89466		
QY 3323	cagccgcggttttttaacccgcgacctctcagcggttgaatgaacagcgctccaccctccag	3382		
DB 89467	CAGCCGGGTTTTTAAACCCCGACCTCTCAGCGCTGTAATGAAACAGCGCTCCACCTCCAG	89526		
QY 3383	tcctggcatccgctgggggcgctgttcttagctagtcgcagttataaaacattgtcatt	3442		
DB 89527	TCCTGGCATCCGCTGGGGGGCGCTTCTTTAGCTAGTGCAGATTAAAAACATTGTCAAT	89586		

RESULT		7			
G22662/c					
LOCUS	483 bp	DNA	STS		31-MAY-1996
DEFINITION	human STS WI-14178, sequence tagged site.				
ACCESSION	G22662				
VERSION	G22662.1	GI:1342988			
KEYWORDS	STS; STS sequence; primer; sequence tagged site.				
SOURCE	human STSS derived from sequences in dbEST and the Unigene collection.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	1 (bases 1 to 483)				
TITLE	Hudson.T				
	Whitehead Institute/MIT Center for Genome Research; Physically				
JOURNAL	Mapped STSS				
COMMENT	Unpublished (1995)				
	Contact: Thomas Hudson				
	Whitehead Institute/MIT Center for Genome Research				



```

RESULT 8
LOCUS HSM801276 1438 bp mRNA PRI 18-FEB-2000
DEFINITION Homo sapiens mRNA; cDNA DKFZp34M2221 (from clone DKFZp34M2221);
partial cds.
ACCESSION AL122052
VERSION AL122052.1 GI:6093247
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1438)
AUTHORS Blum,H., Bauersachs,S., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE Direct Submission
JOURNAL Submitted (15-OCT-1999) MIPS, Am Klopferspitze 18a, D-82152
Martinsried, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by LMU (Ludwig Maximilians University,
Munich/Germany) within the cDNA sequencing consortium of the German
Genome Project.
This clone (DKFZp34M2221) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further
information about the clone and the sequencing project is available
at http://www.mips.biochem.mpg.de/proj/cDNA/.
FEATURES
Location/Qualifiers
1..1438
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp34M2221"
/clone.lib="434 (synonym: htes3). Vector pSport1; host
DH10B; sites NotI + SalI"
/dev_stage="adult"
/tissue_type="testis"
/map="1111.4 CR from top of Chr2 linkage group"
1..694
/gene="DKFZp34M2221"
<1..694
/gene="DKFZp34M2221"
/note="KIAA0793 protein C-terminus"
/codon_start=2
/product="hypothetical protein"
/protein_id="CAB59185.1"
/db_xref="GI:6093248"
/translation="AAQKTVVAASTRLEKEKWMLDNSAIOAAKSGGDTAPALPGRT
VCTRPSPNEVLEQSEDDARGVRSLEGHGQHRANTMHVWYRNTSVSRADHSA
AVENQLSGYLLKFKNSHGKWLWVTFNCLFFYKTHDDYPLASPLLYGSVIPR
EADGTHKDYVFKLOFQSHVYFFRAESKYTFERMEVIOGASSAGRAPSIYQDGPQS
SGLEGVVGKEE"
polyA_signal 1388..1393
polyA_site 1409
BASE COUNT 366 a 426 c 395 g 251 t
ORIGIN
Query Match 8.3%; Score 286.6; DB 9; Length 1438;
Best Local Similarity 68.8%; Pred. No. 5.8e-55;
Matches 426; Conservative 0; Mismatches 184; Indels 9; Gaps 2;
QY 2539 cggcagtcacatcgatgcccaggtctcgccagatgagaagtggttagagac 2598
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 8 CAGAAACAATCGTGGTGGACCCAGCAGCTCGGTGGGAGAAAGAGGATGCTGGAC 67
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2599 atccagatggccattgacctggcggagagagcagcagcccgccctgagttcctggcc 2658
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 68 CTGAACCTCGCGATCCACAGCAGCAAGAGTGGGGTGACACGCGCCCTCCACTGCCAGGC 127
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2659 agca-----gcccccgtgacaacaagtccctctgatgaagcaccgcggtgaccaggag 2712
| | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 128 CGCAGTGTGTGCACATCGTCCCTCCAGATCCCCAACAGAGGTATCTCTG---GAGCAGGAG 184
QY 2713 tcagagatgacctgagcgctcgcgccatcgctgagcgccagcccgcccgccagcgcg 2772
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 185 TCAGAGATGATGCTCGGGGTGTCCGACGTCCCTGGAGGGGATGCCAGCAGCCGGGCC 244
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2773 aaacaatggtgcaagtgctggtgacccacacacacacacacacacacacacacacac 2832
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 245 AACACCACAATGCAGTGTGCTGTACCGGAACACCGAGGTGTCCAGGCGACACACAGT 304
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2833 atcgcatgagaaatcagttgtctgaaacacctgctgagaaattcaaaacacagcaacg 2892
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 305 GCAGCTGTGCGAAGAACACAGCTTTCAGGATATCTGCTAAGAAAGTTCAAAACAGTCTATGCC 364
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2893 tggcagaagctggtggtgtgttcacaaactctgctgtctcttctacaaatcacaccag 2952
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 365 TGGCAGAAGCTCTGGGTGCTCTTACCAACTTCTGTTTGTCTTCTTACAAAACATCATCAG 424
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2953 gacaatcatccctctgcccagctgctctgctgctgctgctgctgctgctgctgctgctgag 3012
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 425 GATGACTACCCACTGGCCAGCCCTCCGCTGCTGGGTACAGCGTGAGCATCCCCAGGGAG 484
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 3013 tccgagaacatccaaaagactacgtgttcaagctgctcactcaagtcacacgtctactac 3072
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 485 GCCGATGGCATACACAAAGACTATGTTTCAAGCTCCAGTCCAAATCCCACTCTACTTC 544
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 3073 ttccagggcgaagcgaagtcacacgttcgaaagtgatggaagtcacgcagtcgaccc 3132
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 545 TTCGGGCTGAGAGCAAGTACACATTTGAAGGTGGATGGAGTGCACGGGGGCCAGC 604
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 3133 agctctgctctgcgacccc 3151
| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 605 AGCTCAGCGGGAGGGGCC 623

RESULT 9
LOCUS AL161896 96183 bp DNA PRI 20-JAN-2001
DEFINITION Human DNA sequence from clone RP11-261P24 on chromosome 13,
complete sequence.
ACCESSION AL161896
VERSION AL161896.16 GI:12330752
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 96183)
AUTHORS Phillimore,B.
TITLE Direct Submission
JOURNAL Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk
COMMENT On Jan 22, 2001 this sequence version replaced gi:12330752.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information
on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/Celegans/wormep/
This sequence
was generated from part of bacterial clone contigs of human
chromosome 13, constructed by the Sanger Centre Chromosome 13
Mapping Group. Further information can be found at

```



```

repeat_region /note="8 copies 6 mer gtgtgt 100% conserved"
38553..39205
/note="LMB3 repeat: matches 5545..6183 of consensus"
39561..39944
/note="MGA repeat: matches 1..426 of consensus"
40827..41086
/note="AluX repeat: matches 1..259 of consensus"
41793..42089
/note="AluSg repeat: matches 1..297 of consensus"
42330..42393
/note="L2 repeat: matches 2627..2701 of consensus"
42692..42857
/note="L1MC2 repeat: matches 5654..5816 of consensus"
42858..43151
/note="AluSg repeat: matches 1..293 of consensus"
43152..43183
/note="L1MC2 repeat: matches 5816..5846 of consensus"
43184..43479

Query Match 7.4%; Score 254; DB 9; Length 96183;
Best Local Similarity 100.0%; Pred. No. 1.3e-47;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1210 caGtcagcagacacagccttacatttgagaaagtgcgaatctccaggggcccag 1269
|||||
Db 95665 CAGTCTCAGCAGACACAGCCGCTTACATTTGGAGAGGTGCCGAATCTCCAGGGGGCCAG 95724
|||||

QY 1270 agTtcgCGgaggaagaaacacgaagtttccgCGgggagcgcgggtcgaccccgagc 1329
|||||
Db 95725 AGCTGCCGGCAGGAGGAAGAACCGAAGTTTCGCCGGGAGCGGGGTGCACCCCGAGC 95784
|||||

QY 1330 cctgcgcgaggaagaccccgCGgggttaacaagcagcgagcagcgcgcgcgcgc 1389
|||||
Db 95785 CCTGCCGGAGGAGAACCGCGGGTAAACAGCAGCGCGGCGGCTCGCGCGCC 95844
|||||

QY 1390 acgagaaagagagaggttgaagatagagaccagacagataaacctcacccccc 1449
|||||
Db 95845 ACGGAGGAAGAGGAGGAGGTGTTAAGATAGGACCCAGCAGAGATAACCTCAGCCCGCC 95904
|||||

QY 1450 cagccaagcacag 1463
|||||
Db 95905 CAGCAGACACAGG 95918

RESULT 10
AC022669/c
LOCUS AC022669 152053 bp DNA HTG 04-MAY-2001
DEFINITION Homo sapiens chromosome 15 clone RP11-96B23 map 15, WORKING DRAFT
SEQUENCE, 33 unordered pieces.
AC022669
VERSION AC022669.4 GI:9966291
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 152053)
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
Boquslavkiy,L., Bouhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepe,X., Collangelo,M., Collins,S., Collymore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,

```

TITLE  
JOURNAL  
COMMENT

Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Rhmann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A., Vassiliev,H., Viel,R., Vo.A., Wu,X., Wyman,D., Ye.W.J., Zimmer,A. and Zody,M.

Direct Submission  
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 4, 2000 this sequence version replaced gi:729803.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L4938  
Center clone name: 96\_B23  
----- Summary Statistics  
Sequencing vector: M13; M77815; 93% of reads  
Sequencing vector: Plasmid; n/a: 0.0% of reads  
7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 136451 bases at least Q40  
Consensus quality: 143612 bases at least Q30  
Consensus quality: 146365 bases at least Q20  
Insert size: 158000; agarose-fp  
Insert size: 148853; sum-of contigs  
Quality coverage: 3.4 in Q20 bases; agarose-fp  
Quality cov.  
NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 784: contig of 784 bp in length  
885 884: gap of 100 bp  
885 1956: contig of 1072 bp in length  
1957 2056: gap of 100 bp  
2057 3113: contig of 1057 bp in length  
3114 3213: gap of 100 bp  
3214 4817: contig of 1604 bp in length  
4818 4917: gap of 100 bp  
4918 6582: contig of 1665 bp in length  
6583 6682: gap of 100 bp  
6683 8059: contig of 1377 bp in length  
8060 8159: gap of 100 bp  
8160 22369: contig of 14210 bp in length  
22370 22469: gap of 100 bp  
22470 24974: contig of 2505 bp in length  
24975 25074: gap of 100 bp  
25075 27206: contig of 2132 bp in length  
27207 27306: gap of 100 bp  
27307 29420: contig of 2114 bp in length  
29421 29520: gap of 100 bp  
29521 32000: contig of 2480 bp in length  
32001 32100: gap of 100 bp  
32101 36330: contig of 4230 bp in length  
36331 36430: gap of 100 bp  
36431 39067: contig of 2637 bp in length  
39068 39167: gap of 100 bp  
39168 42564: contig of 3397 bp in length  
42565 42664: gap of 100 bp  
42665 46664: contig of 4000 bp in length  
46665 46764: gap of 100 bp  
46765 51047: contig of 4283 bp in length  
51048 51147: gap of 100 bp



requests: clonerequest@sanger.ac.uk  
 On Nov 8, 2000 this sequence version replaced gi:11061750.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information  
 on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/Projects/C.elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 13, constructed by the Sanger Centre Chromosome 13  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr13  
 This sequence is the entire insert of clone RP11-10G5 The true left  
 end of clone RP11-573N10 is at 109104 in this sequence. The true  
 right end of clone RP11-72J7 is at 95284 in this sequence. This  
 sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. RP11-10G5 is from the  
 library RP11-11.1 constructed by the group of Pieter de Jong. For  
 further details see  
 http://www.chori.org/bacpac/home.htm  
 VECTOR: pBACE3.6.

## FEATURES

source	Location/Qualifiers	misc_feature	/note="63 copies 2 mer ca 88% conserved"
	1. .170916	41206. .41622	/note="match: GSS: Em:AQ184690"
	/organism="Homo sapiens"	42604. .43109	/note="match: GSS: Em:AQ573113"
	/db_xref="taxon:9606"	43382. .43429	/note="24 copies 2 mer ta 93% conserved"
	/chromosome="13"	complement(46659. .47093)	/note="match: GSS: Em:AQ805998"
	/clone="RP11-10G5"	46717. .46754	/note="19 copies 2 mer gt 100% conserved"
	/clone_lib="RPC1-11.1"	53170. .53274	/note="MER34 repeat: matches 8. .115 of consensus"
	384. .873	53278. .53412	/note="MER74B repeat: matches 50. .180 of consensus"
repeat_region	/note="MER31B repeat: matches 1. .481 of consensus"	55996. .56438	/note="MER83 repeat: matches 1. .448 of consensus"
repeat_region	1374. .1407	59536. .60314	/note="MER21B repeat: matches 63. .794 of consensus"
repeat_region	/note="17 copies 2 mer tt 85% conserved"	60344. .60392	/note="MER21B repeat: matches 741. .790 of consensus"
misc_feature	5470. .5533	60825. .61336	/note="MER74B repeat: matches 61. .621 of consensus"
repeat_region	/note="32 copies 2 mer tt 70% conserved"	61616. .61665	/note="MER61B repeat: matches 377. .425 of consensus"
repeat_region	6352. .6714	61667. .62339	/note="LFR9 repeat: matches 1. .625 of consensus"
repeat_region	/note="match: GSS: Em:AQ698072"	62342. .62683	/note="MER61C repeat: matches 1. .384 of consensus"
repeat_region	8498. .8665	63996. .64135	/note="MLT2E repeat: matches 1. .149 of consensus"
repeat_region	/note="TIGGER1 repeat: matches 2080. .2238 of consensus"	64486. .64768	/note="MLT2E repeat: matches 149. .395 of consensus"
repeat_region	8670. .8888	65150. .65369	/note="LJMA5A repeat: matches 6053. .6291 of consensus"
repeat_region	/note="MER75B repeat: matches 18. .239 of consensus"	65371. .66037	/note="LPR12 repeat: matches 1. .647 of consensus"
repeat_region	13335. .13686	66038. .66088	/note="LJMA5A repeat: matches 6037. .6091 of consensus"
repeat_region	13996. .14238	66076. .66299	/note="LJMA5A repeat: matches 5827. .6047 of consensus"
repeat_region	/note="176 copies 2 mer tg 69% conserved"	67351. .67427	/note="MLT1A2 repeat: matches 173. .250 of consensus"
repeat_region	/note="LIPB3 repeat: matches 5894. .6150 of consensus"	67428. .67487	/note="30 copies 2 mer ag 88% conserved"
repeat_region	15610. .15639	67492. .67660	/note="MLT1A2 repeat: matches 5. .174 of consensus"
repeat_region	/note="15 copies 2 mer ta 86% conserved"	68268. .68810	/note="MER68A repeat: matches 1. .561 of consensus"
repeat_region	complement(17881. .18076)	complement(73998. .74451)	/note="match: GSS: Em:AQ695161"
repeat_region	/note="match: GSS: Em:AQ776608"	74143. .74773	/note="LTR39 repeat: matches 172. .794 of consensus"
repeat_region	18339. .18804	75063. .75233	/note="LTR39 repeat: matches 1. .172 of consensus"
repeat_region	19891. .20105	79181. .79291	/note="MLT1A2 repeat: matches 251. .374 of consensus"
repeat_region	/note="LFR7 repeat: matches 1. .448 of consensus"	79699. .80013	/note="MLT1A2 repeat: matches 1. .251 of consensus"
repeat_region	/note="MER57-internal repeat: matches 7171. .7387 of consensus"	80095. .80463	/note="THE1C repeat: matches 1. .369 of consensus"
repeat_region	20131	83250. .83460	/note="Tigger3b repeat: matches 1. .210 of consensus"
repeat_region	/note="Tandem repeat. Forced join. Gap size estimated to be approximately 260bp by restriction digest data."	83757. .84298	/note="Tigger3b repeat: matches 210. .750 of consensus"
repeat_region	29106. .29251	84581. .84607	/note="Tigger3b repeat: matches 750. .775 of consensus"
repeat_region	31944. .32589	84608. .84963	/note="THE1C repeat: matches 4. .371 of consensus"
repeat_region	/note="LFR12 repeat: matches 1. .671 of consensus"	84964. .85004	/note="Tigger3b repeat: matches 775. .825 of consensus"
repeat_region	32590. .33323	85677. .86088	/note="Tigger3b repeat: matches 823. .1231 of consensus"
repeat_region	/note="PTR5 repeat: matches 743. .1478 of consensus"		
repeat_region	38726. .38851		

```

misc_feature      complement(89463...90009)
                  /note="match: GSS: Em:AQ013482"
misc_feature      complement(89490...90015)
                  /note="match: GSS: Em:AQ383425"
misc_feature      complement(89526...90008)
                  /note="match: GSS: Em:B82154"
misc_feature      complement(93764...94093)
                  /note="match: GSS: Em:AQ594661"
misc_feature      98631...100473
                  /note="CpG island"
                  /evidence=not_experimental
misc_feature      99666...99802
                  /note="match: GSS: Em:AQ938343"
misc_feature      99694...99750
                  /note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
repeat_region     100689...100712
                  /note="12 copies 2 mer tg 100% conserved"
misc_feature      101250...101669
                  /note="match: GSS: Em:AQ008502"
misc_feature      101284...101479
                  /note="match: GSS: Em:AQ000309"
repeat_region     105297...105359
                  /note="LIMEC repeat: matches 1638. 1710 of consensus"
misc_feature      108760...109211
                  /note="match: GSS: Em:B66665"
misc_feature      108771...109149
                  /note="match: GSS: Em:AQ076004"
misc_feature      108818...109537
                  /note="match: GSS: Em:AQ307265"
repeat_region     109801...110016
                  /note="108 copies 2 mer at 58% conserved"
misc_feature      117543...118068
                  /note="match: GSS: Em:AQ678189"
repeat_region     118206...118247

Query Match      5.8%; Score 199.2; DB 9; Length 170916;
Best Local Similarity 98.5%; Pred. No. 4.3e-35;
Matches 201; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 21 tggagatattcttaagccgtttcatcattgggagaataagcagagccgacccagg 80
Db 169701 TGCAGATATTCTCTAAGCCGTTTCATCATGGGAGAAATAGAGCAGAGCGACCCAGG 169760

QY 81 atcacgactggggccgcaaatcgggacgacgtacacgttgaaactgacagaagcc 140
Db 169761 ATCAGGACTGGGGCCCGGAAATTCGGGGATCAGTACCTTGGAACTGGACAGAGCC 169820

QY 141 gcccccaacaccttcagaaaactcgtgtccatcaataacacacacacacacacac 200
Db 169821 GCCCCCAACACCTTCAGGAAACCTCGTGTCATCAATCAAAATCCAGATGCTGGATGACACCCA 169880

QY 201 ggaggcaattgaagttccacaag 224
Db 169881 GGAGGCATTTGAAGTTCAGTAAG 169904

RESULT 12
AF156225 LOCUS      2595 bp mRNA PRI 04-JUL-1999
DEFINITION Homo sapiens protein 4.1 (EPB4.1) mRNA, complete cds.
ACCESSION AF156225
VERSION AF156225.1 GI:5353737
KEYWORDS      human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2595)
AUTHORS Huang,S.C., Wang,C., Lichtenauer,U., Vortmeyer,A. and Zhuang,Z.
TITLE Sequence of protein 4.1 from a human neuroblastoma cell line: LAN5
JOURNAL Unpublished

```

```

REFERENCE 2 (bases 1 to 2595)
AUTHORS Huang,S.C., Wang,C., Lichtenauer,U., Vortmeyer,A. and Zhuang,Z.
TITLE Direct Submission
JOURNAL Submitted (03-JUN-1999) NCI, NIH, 9000 Rockville Pike, Bethesda, MD
20892, USA
FEATURES
    source
    1..2595
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /chromosome="1"
        /map="lp36"
        /cell_line="LAN5"
        /cell_type="neuroblastoma"
    1..2595
        /gene="EPB4.1"
    1..2595
        /gene="EPB4.1"
        /codon_start=1
        /product="protein 4.1"
        /protein_id="AAD4222.1"
        /db_xref="GI:5353738"
        /translation="MTETSLVTEAENSQHQKEEGEALNSQOEPQOPEESCOTAAE
        GDNWCHKLKASNGDTTHEDTLKNERISENGLRFLSFSLKRPKSQVSEEGKEV
        ESKKEGGQKEIEFGTSLDEEILKAPIAAPELKTDPDLSLHSLSAETQPAQE
        ELREDPFKEIGEGLECKIEVKESPAQAEELKASQKPIRKHNNHCKVSLLD
        DTVYECVEKHAKGQDLKRVCEHLNLEEDYFGLAIWNTTSTKTLWLSAKELKKQVR
        GVPWNFTNVKFPDPAOLTEDITRYLCLQLRQDIVAGRLPCSTATLALLGSYTIQ
        SELGDYDELHGVYVSDFKLAPNOTKEELKVMELHKSYSRMTPAQADLEFLFNKK
        LRMVGLDHAKDLEVDIILGVCSGLLYIKDKLRNRPWPVKYALISYRSRFFIK
        IRPGEQYEYESTIGFKLPSYRAAKLWKVCVHHFTSTDTIPKSKFLAGSKFR
        YSGTQAOTQASALIDRPAPHPERTASRSLDGAADAAVSDRSPRTSAPAITQ
        GOVAEGGVLDASAKTVVPAQKQETVKAQKEDPEQAEPEPEAWKVEKTHIEVT
        VPTNSGDQTKLAEKTEDLIRMKKKRRLDGENIYIRHSSLMLELDLKSQREIEKHH
        ASISLKNFMESVPEPRSEWDKRLSTHSPPTLNINGQIPIGEGPPLVKTQTVTIS
        DNANAVSEIPNDPFIHVTETKTIITYAAQDDNSGDLDPGLLVLLTAQITITSTPST
        TTTQITTKVKGISSETRIEKRIVITGDADIDHDQVLVQAIKEAKEQHPDMSVTKVVVH
        QETEIADE"
BASE COUNT      881 a 563 c 601 g 550 t
ORIGIN
Query Match      5.6%; Score 194; DB 9; Length 2595;
Best Local Similarity 52.9%; Pred. No. 8.9e-34;
Matches 491; Conservative 0; Mismatches 425; Indels 12; Gaps 3;

QY 165 cgtgtccatcaaaatccagatgctggatgacacccacagagcatttgaaagttccacaag 224
Db 627 CATGCACCTGCAAGGTTCTTGTGGATGACAGTTTATGAATGTTGTGGAGAAACA 686

QY 225 agtcctctgggaaggtgctgctggatgcagttgcaaccacacacacacacacacac 284
Db 687 TGCTAAGGGACAGATTGCTTAAACGAGTATGTGACCATCTCAATCTTTTGGAGAAGA 746

QY 285 ctatttggcctcgagttctctgacacaaaagatcacggtgtggtgctgacatcctctaa 344
Db 747 CTATTTTGGTCTAGCCATTTGGGATAACACACCTCTAAGACATGCGTGGATTCCGCAA 806

QY 345 acccattgtgaaacagattagaagggccaaagcacgttgttgtaagtttgtggaatt 404
Db 807 AGAAATAAAAAACAGGTTCTGTGG---TGTCCTTGGAAATTTTACATTTAATGTAAAGTT 863

QY 405 ctcttcgcctgcacacacacacacacacacacacacacacacacacacacacacacac 464
Db 864 TTATCCACCTGACCCAGCAGTAAACAGAGACATACAGATATATTTATGCTCTCA 923

QY 465 ggtgaagcaggaacttggtctcaagcaggttgacgtgtaatgacacacgacgactctctt 524
Db 924 GCTTCGCGAGGACATAGTTGCGAGCAGCTGCGCTCTCTCTTTGCAACCTTAGCATATT 983

QY 525 gattcacacatgtcaactcagattggggattgagccttgacagccttgacagagacca 584
Db 984 AGGTTCTTACCATCCAGTCTGAACCTGGGAGACTAGCACCAGAACTCCATGCGGTGGA 1043

```

**TITLE**  
**JOURNAL**





```

CDS
799..2565
/note="4.1 protein"
/codon_start=1
/protein_id="AAA35795.1"
/db_xref="GI:182076"
/translation="MCKKVSLLDDTVYECVVEKHAKGQDLLKRVCEHLNLEEDYFGL
AIDNATSKTWDLSAKEIKKQVRGVPWNFTFNKFPDPAQLTEDITRYILCLLRQ
DIVAGRLPCSFATLALIGSYTIOSELGDYDPELHGVYVSDFKLAIPNQTKELEKVM
LHKSYSMTPAQADLEFLNAKLSMYGVDLHRAKDLGVDIILGVCSSGLLVYKDL
RINRFPWPKVLKISYKRSFFIKIRPEQOEQESTIGFKLPSYRAAKKLWKVCVEHHT
FFRLTSDTIPKSKFLALGSKFRYSGRTOAQASALIDRPAPHFERTAKRSRL
DGAAYDSADSPRPTSAIPATOGVAEGVLDASAKKTVVPAQKERTVKAERVKDE
PPQAEPEPEAWKREKRLDGENIIRHNLMLDLSQEEIKKHHSISLKNK
MESVPRSEMDKRLSTHSPFTLNINGOIPTEGPPVLKQTQVITSDNANAVKSEI
PKTDPVIVITETIYEAQVTKGGISETRIEKRIIVITGDADIDHDQVLOAIKEAK
EQHPDMSVTKVVVHQETEADE"
2016..2017
/variation
/note="duplication insertion of nucleotides 2217-2585
causing elliptocytosis [2]"
2017..2256
/variation
/note="deletion causing elliptocytosis [2]"
BASE COUNT 932 a 620 c 657 g 658 t
ORIGIN 218 bp upstream of PvuII site.

Query Match 5.6%; Score 192.4; DB 9; Length 2867;
Best Local Similarity 52.8%; Pred. No. 2.1e-33;
Matches 490; Conservative 0; Mismatches 426; Indels 12; Gaps 3;

Qy 165 cgtgtccatacaaatccagatgctgtagacacccaggagcatttgaagttccacaaag 224
Db 165 CATGCTGCAAGGTTCTTGTGGATGACACAGTTTATGAATGCTGTGGAGAAACA 857
Qy 225 agctccctgggaagtgctgctggatgagcttgcaaccacccacacccctcggtggaagtga 284
Db 225 AGCTCCCTGGGAAGTGGCTTAACAGGATGATGAGCACTCAATCTTTTGGAGGAAGA 917
Qy 285 ctatttggcctgagtttctgctgacacaaagatcacggtggtggtggtatctctctaa 344
Db 285 CTATTTTGGTCTGCCATTGGGATAACGCAACCTCTAAGACATGGCTGGATTCCGCCAA 977
Qy 345 accattgtgaacacagattagaagccaaagcagctgtgtgttaagtttgtgtggaatt 404
Db 345 ACAAATTAAGCAAGCAGGTGCTGTG---TGTCCTTGGAAATTTTACATTTAATGTAAGATT 1034
Qy 405 ctttcgcctgacacacacacactccaagaagactcacaggtagctgttcgcgtgcga 464
Db 405 TATCCACCTGACCCAGCACAGTTAAACAGAGACATAACAGATATTATTTATGCTCTCA 1094
Qy 465 ggtgaacagagacttggtcgaagcaggttgacgtgtaataagacaccagcagctctctt 524
Db 465 GCTTTCGGCAGGACATAGTTGCAGGACGCTCTGCCCTGTCTCTTTCACACCTTAGCATTT 1154
Qy 525 gatttcacacattgtcaatctgagattgggatttggatgaaagccttggacagagaca 584
Db 525 AGGTTCTTACACCATTCAGTGTGAACCTGGGAGACTACGCCAACATCCATGCCGTGGA 1214
Qy 585 cttagcaaaaaataataacataacc-----tcagcaagacgcactagagacaaaatcgt 638
Db 585 TATGTGTAGTATTTTAACTGGCCCCGAATCACACCAAGGAACCTTGAAGAGAGGTCAT 1274
Qy 639 ggaatttcacataaccacattggacaaacaccagcagagaatcagatttccagctcctaga 698
Db 639 GGAACCTGCATTAAGTCATACAGGTCATGACTCCAGCTCAGCTGACCTGAGGTTCTTGA 1334
Qy 699 gattgcctcggtcgtagatgtagaattccggttgcaccccgccagagacagggagg 758
Db 699 GAATGCCAAAAGTGTCTATGTATGGAGTTGATCTTCATAAAGCAAGAGACTTGGAGG 1394
Qy 759 caccgaagatacaatctgcccgtgtgcaacacgggaattcttagtgtttcagggttcaactaa 818
Db 759 AGTAGATATCATCTAGTGTCTGCTAGTGGCTCTCTGCTTTTACAAAGATAAGCTGAG 1454

```

```

Qy 819 gataaatgccttcaactgcccgaaggtgcggaagctgagcttcaagaggagcgctttct 878
Db 1455 AATTAACCGCTTCCCTTGGCCCAAAGTCTGAAGATTTCTTATTAACGCTAGTAGCTTTT 1514
Qy 879 catcaagctcgcgcagatgccaatagtcgtagcagacacacattggaattccctgatg 938
Db 1515 CATCAAGATTCCGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1574
Qy 939 cagtcgggatttctgcaagctcctctggaataatctgtgtgaacatcatgctcctctt 998
Db 1575 CAGTTACCGAGCAGCTAAGAAATTTATGAAAGTCTGTGTAGAACATCACACGTTTTCAG 1634
Qy 999 acttttgaagagcccaacaaagcccaagcccgctcctctttagccggggggtcatcatt 1058
Db 1635 A---TTGACATCTACAGACACCATTTCCCAAAAGCAAAATTTCTTCGCTAGATCCAAAT 1691
Qy 1059 tcggttcagtggtcgactcagaagcag 1086
Db 1692 TCGATACAGTGGCGGACTCAAGCTCAG 1719

RESULT 15
HUMEMP41 3043 bp mRNA PRI 03-MAR-1994
LOCUS Homo sapiens erythroid membrane protein 4.1 mRNA, complete cds.
DEFINITION M61733
ACCESSION M61733.1 GI:182081
VERSION 1
KEYWORDS erythroid membrane protein 4.1.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 3043)
AUTHORS Conboy,J.G., Chan,J.Y.C., Chasis,J.A., Kan,Y.W. and Mohandas,N.
TITLE Tissue- and development-specific alternative RNA splicing regulates
expression of multiple isoforms of Erythroid membrane protein 4.1
J. Biol. Chem. 266, 8273-8280 (1991)
JOURNAL 91217063
MEDLINE
FEATURES
source Location/Qualifiers
1..3043
/organism="Homo sapiens"
/db_xref="taxon:9606"
816..2741
/codon_start=1
/product="erythroid membrane protein 4.1"
/protein_id="AAA35797.1"
/db_xref="GI:182082"
/translation="MCKKVSLLDDTVYECVVEKHAKGQDLLKRVCEHLNLEEDYFGL
AIDNATSKTWDLSAKEIKKQVRGVPWNFTFNKFPDPAQLTEDITRYILCLLRQ
DIVAGRLPCSFATLALIGSYTIOSELGDYDPELHGVYVSDFKLAIPNQTKELEKVM
LHKSYSMTPAQADLEFLNAKLSMYGVDLHRAKDLGVDIILGVCSSGLLVYKDL
RINRFPWPKVLKISYKRSFFIKIRPEQOEQESTIGFKLPSYRAAKKLWKVCVEHHT
FFRLTSDTIPKSKFLALGSKFRYSGRTOAQASALIDRPAPHFERTAKRSRL
DGAAYDSADSPRPTSAIPATOGVAEGVLDASAKKTVVPAQKERTVKAERVKDE
PPQAEPEPEAWKREKRLDGENIIRHNLMLDLSQEEIKKHHSISLKNK
MESVPRSEMDKRLSTHSPFTLNINGOIPTEGPPVLKQTQVITSDNANAVKSEI
PKTDPVIVITETIYEAQVTKGGISETRIEKRIIVITGDADIDHDQVLOAIKEAKEQHPD
MSVTKVVVHQETEADE"
BASE COUNT 997 a 673 c 690 g 683 t
ORIGIN

```

```

Query Match 5.6%; Score 192.4; DB 9; Length 3043;
Best Local Similarity 52.8%; Pred. No. 2.1e-33;
Matches 490; Conservative 0; Mismatches 426; Indels 12; Gaps 3;

Qy 165 cgtgtccatacaaatccagatgctgtagacacccaggagcatttgaagttccacaaag 224
Db 815 CATGCTGCAAGGTTCTTGTGGATGACACAGTTTATGAATGCTGTGTGGAGAAACA 874
Qy 225 agctccctgggaagtgctgctgtagacacccacacacacacacacacacacacacacac 284

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:55:11 ; Search time 36.72 Seconds  
(without alignments)  
55.768 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_764\_854  
Perfect score: 468  
Sequence: 1 GSKLSKGLQRMFFLN.....IVAASRSEMEKWKVEDIQMA 91

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	84.5	18.1	394	2	US-08-839-581A-4
2	84.5	18.1	394	4	US-09-023-591A-4
3	80.5	17.2	814	1	US-08-318-831-2
4	79.5	17.0	1244	3	US-09-356-952-7
5	75.5	16.1	399	2	US-08-839-581A-2
6	75.5	16.1	399	4	US-09-023-591A-2
7	71.5	15.3	397	4	US-09-284-033-4
8	71.5	15.3	397	4	US-08-729-834B-4
9	71.5	15.2	127	4	US-09-284-033-6
10	71.5	15.2	127	4	US-08-729-834B-6
11	71.5	15.2	398	2	US-08-839-581A-32
12	71.5	15.2	398	4	US-09-023-591A-32
13	70.5	15.1	126	4	US-09-284-033-5
14	70.5	15.1	126	4	US-08-729-834B-5
15	70.5	15.1	399	4	US-09-284-033-2
16	70.5	15.1	399	4	US-08-729-834B-2
17	69.5	14.9	537	2	US-08-633-879C-2
18	67.5	14.4	689	1	US-08-221-817-18
19	67.5	14.4	689	1	US-08-454-439-18
20	67.5	14.4	689	5	PCT-US94-10487-18
21	66.5	14.2	136	2	US-08-943-208-2
22	63	13.5	1297	2	US-08-290-731C-4
23	61	13.0	3077	6	5223423-2
24	60	12.8	496	2	US-08-463-418-2
25	59.5	12.7	363	4	US-08-981-234B-2
26	59.5	12.7	527	4	US-08-592-126-145
27	59.5	12.7	527	2	US-08-687-080-48

28	59.5	12.7	535	2	US-08-633-879C-4	Sequence 4, Appli
29	58.5	12.5	668	4	US-09-436-605-2	Sequence 2, Appli
30	58.5	12.5	688	1	US-08-221-817-19	Sequence 19, Appl
31	58.5	12.5	688	1	US-08-454-439-19	Sequence 19, Appl
32	58.5	12.5	688	5	PCT-US94-10487-19	Sequence 19, Appl
33	57.5	12.3	266	3	US-08-815-809-5	Sequence 5, Appli
34	57.5	12.3	280	2	US-08-816-155B-43	Sequence 43, Appl
35	57.5	12.3	280	4	US-09-079-587-43	Sequence 43, Appl
36	57.5	12.3	374	3	US-08-878-801-2	Sequence 2, Appli
37	57.5	12.3	502	1	US-08-484-840-3	Sequence 3, Appli
38	57.5	12.3	502	1	US-08-483-094-3	Sequence 3, Appli
39	57.5	12.3	646	4	US-09-232-200-59	Sequence 59, Appl
40	57.5	12.3	646	4	US-09-232-197-59	Sequence 59, Appl
41	57	12.2	298	2	US-08-838-543-5	Sequence 5, Appli
42	57	12.2	796	2	US-08-817-900-2	Sequence 2, Appli
43	57	12.2	796	4	US-09-236-645-2	Sequence 2, Appli
44	56.5	12.1	659	1	US-08-136-277-3	Sequence 3, Appli
45	56.5	12.1	659	2	US-08-479-403-3	Sequence 3, Appli

ALIGNMENTS

RESULT 1  
US-08-839-581A-4  
; Sequence 4, Application US/08839581A  
; Patent No. 5958705  
; GENERAL INFORMATION:  
; APPLICANT: Staunton, Donald E.  
; APPLICANT: Lipsky, Brian P.  
; TITLE OF INVENTION: Cytoplasmic Modulators of Integrin  
; TITLE OF INVENTION: Binding/Signaling  
; NUMBER OF SEQUENCES: 32  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
; STREET: 233 South Wacker, 6300 Sears Tower  
; CITY: Chicago  
; STATE: Illinois  
; COUNTRY: USA  
; ZIP: 60606  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/839,581A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Young J. Suh  
; REGISTRATION NUMBER: P-41,337  
; REFERENCE/DOCKET NUMBER: 27866/33886  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-474-6300  
; TELEFAX: 312-474-0448  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 394 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-839-581A-4

Query Match 18.1%; Score 84.5; DB 2; Length 394;  
Best Local Similarity 25.7%; Pred. No. 0.0037;  
Matches 29; Conservative 22; Mismatches 33; Indels 29; Gaps. 5;  
QY 1 GSKLSKSG--KGLQRMFFLNVDVLLYTSRGLTASNQFKVHGOLPLGYGMTIESEDEWGV 58  
DB 264 GWLLKLGKRVKTRWRRIITDNCILY----FEFTTKRGIIPLENLSQVKVDDP-KK 318



```
Query Match 17.0%; Score 79.5; DB 3; Length 1244;
Best Local Similarity 26.9%; Pred. No. 0.084;
Matches 29; Conservative 14; Mismatches 38; Indels 27; Gaps 3;

QY 1 GSLSKLSGKGLQORFFLENDVLLYTSRGLTASNQFKVHGQPLVGMRI---ESEDENG 57
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 477 GRGLSLKSKKEGRCQFLFSKHLIICITRGSGLKHLTRNGVISLTDCTLLDDPENDDG 536
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 58 VPHCLTLRGQO-----SIIVAASSRSSEMEKWEEDI 88
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 537 -----KGQVDHLDFKIWEKDPSPFTVILVASSRQKRAWTSDI 577
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 5
US-08-839-581A-2
; Sequence 2, Application US/08839581A
; Patent No. 5958705
; GENERAL INFORMATION:
; APPLICANT: Staunton, Donald E.
; APPLICANT: Lipsky, Brian P.
; TITLE OF INVENTION: Cytoplasmic Modulators of Integrin
; TITLE OF INVENTION: Binding/Signaling
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,581
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Young J. Suh
; REGISTRATION NUMBER: P-41,337
; REFERENCE/DOCKET NUMBER: 27866/33886
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; APPLICATION NUMBER: US/08/839,581A
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Young J. Suh
; REGISTRATION NUMBER: P-41,337
; REFERENCE/DOCKET NUMBER: 27866/33886
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-839-581A-2

Query Match 16.1%; Score 75.5; DB 2; Length 399;
Best Local Similarity 24.6%; Pred. No. 0.061;
Matches 28; Conservative 21; Mismatches 38; Indels 27; Gaps 5;

QY 1 GSLSKLSG--KGLQORFFLENDVLLYTSRGLTASNQFKVHGQPLVGMRIESED---- 54
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 264 GWLLKLGGRVTKRWRFILTDNCLYY----FEYTTDKPRGIIPLENLSIREVDPRKP 319
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 55 ---EWGVPH-----CLT-----LRGQRSIIVAASSRSSEMEKWEEDIQMA 91
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 320 NCFELYIPNNKGOLIKACKTEADGRVVEGNHMYRISAPTQEEKDEWIKSIQAA 373
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 6
US-09-523-591A-2
; Sequence 2, Application US/09023591A
; Patent No. 6210914
```

```
; GENERAL INFORMATION:
; APPLICANT: Staunton, Donald E.
; APPLICANT: Lipsky, Brian P.
; TITLE OF INVENTION: Cytoplasmic Modulators of Integrin
; TITLE OF INVENTION: Binding/Signaling
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/839,581
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Young J. Suh
; REGISTRATION NUMBER: P-41,337
; REFERENCE/DOCKET NUMBER: 27866/33886
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 399 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-023-591A-2

Query Match 16.1%; Score 75.5; DB 4; Length 399;
Best Local Similarity 24.6%; Pred. No. 0.061;
Matches 28; Conservative 21; Mismatches 38; Indels 27; Gaps 5;

QY 1 GSLSKLSG--KGLQORFFLENDVLLYTSRGLTASNQFKVHGQPLVGMRIESED---- 54
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 264 GWLLKLGGRVTKRWRFILTDNCLYY----FEYTTDKPRGIIPLENLSIREVDPRKP 319
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 55 ---EWGVPH-----CLT-----LRGQRSIIVAASSRSSEMEKWEEDIQMA 91
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 320 NCFELYIPNNKGOLIKACKTEADGRVVEGNHMYRISAPTQEEKDEWIKSIQAA 373
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

RESULT 7
US-09-284-033-4
; Sequence 4, Application US/09284033
; Patent No. 6194173
; GENERAL INFORMATION:
; APPLICANT: Czech, Michael P. and Klarlund, Jes K.
; TITLE OF INVENTION: BINDING PROTEINS FOR PHOSPHOINOSITIDES, GRP1 OR GENERAL REC
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/284,033  
FILING DATE: 1999-04-06  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN: 08/729,834  
FILING DATE: 07 OCTOBER 1996  
APPLICATION NUMBER: PCT/US97/18152  
FILING DATE: 1997-10-07  
ATTORNEY/AGENT INFORMATION:  
NAME: MANDRAGOURAS, AMY E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: UMM-018CPUS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 397 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-09-284-033-4

Query Match 15.3%; Score 71.5; DB 4; Length 397;  
Best Local Similarity 24.3%; Pred. No. 0.21;  
Matches 28; Conservative 17; Mismatches 41; Indels 29; Gaps 4;  
QY 1 GSLSKLSG---KGLQRMFLFNDVLYTSRGLTASNQFKVHGOLPLYGMTIESEDEWG 57  
Db 265 GWLLKGGGRVKTWRWFILTDNCLY-----FEYTTDKPRGIIPLENLSIREVEDS-K 319  
QY 58 VPHCLTL-----RGQROSIIVAASSRSEMEKWKVEDIQMA 91  
Db 320 KPNCFELYIPDNKDQVIKACKTEADGRWGNHTVYRISAPTPEKEEWIKCIAA 374

RESULT 8  
US-08-729-834B-4  
Sequence 4, Application US/08729834B  
Patent No. 6221841  
GENERAL INFORMATION:  
APPLICANT: Czech, Michael P.  
APPLICANT: Klarlund, Jes K.  
TITLE OF INVENTION: General Receptors for Phosphoinositides  
TITLE OF INVENTION: and Uses Related Thereto  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/729,834B  
FILING DATE: October 7, 1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Amy E. Mandragouras  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: UMM-018  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941

INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 397 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-729-834B-4

Query Match 15.3%; Score 71.5; DB 4; Length 397;  
Best Local Similarity 24.3%; Pred. No. 0.21;  
Matches 28; Conservative 17; Mismatches 41; Indels 29; Gaps 4;  
QY 1 GSLSKLSG---KGLQRMFLFNDVLYTSRGLTASNQFKVHGOLPLYGMTIESEDEWG 57  
Db 265 GWLLKGGGRVKTWRWFILTDNCLY-----FEYTTDKPRGIIPLENLSIREVEDS-K 319  
QY 58 VPHCLTL-----RGQROSIIVAASSRSEMEKWKVEDIQMA 91  
Db 320 KPNCFELYIPDNKDQVIKACKTEADGRWGNHTVYRISAPTPEKEEWIKCIAA 374

RESULT 9  
US-09-284-033-6  
Sequence 6, Application US/09284033  
Patent No. 6194173  
GENERAL INFORMATION:  
APPLICANT: Czech, Michael P. and Klarlund, Jes K.  
TITLE OF INVENTION: BINDING PROTEINS FOR PHOSPHOINOSITIDES, GRP1 OR GENERAL REC  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD, LLP  
STREET: 28 STATE STREET  
CITY: BOSTON  
STATE: MASSACHUSETTS  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/284,033  
FILING DATE: 1999-04-06  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: USSN: 08/729,834  
FILING DATE: 07 OCTOBER 1996  
APPLICATION NUMBER: PCT/US97/18152  
FILING DATE: 1997-10-07  
ATTORNEY/AGENT INFORMATION:  
NAME: MANDRAGOURAS, AMY E.  
REGISTRATION NUMBER: 36,207  
REFERENCE/DOCKET NUMBER: UMM-018CPUS  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 127 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-09-284-033-6

Query Match 15.2%; Score 71; DB 4; Length 127;  
Best Local Similarity 24.1%; Pred. No. 0.051;  
Matches 28; Conservative 17; Mismatches 41; Indels 30; Gaps 4;

TITLE OF INVENTION: Binding/Signaling  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 333 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA

ZIP: 60808  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, v1.0  
CURRENT APPLICATION DATA: US/08/839,581A  
APPLICATION NUMBER: US/08/839,581A  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Young J. Suh  
REGISTRATION NUMBER: P-41,337  
REFERENCE/DOCKET NUMBER: 2786636/3381  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
INFORMATION FOR SFO ID NO: 32

SEQUENCE CHARACTERISTICS:  
LENGTH: 398 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
-839-581A-32

```

      15.2%; Score 71; DB 2; Length 398;
      at Local Similarity 24.1%; Pred. No. 0.24;
      28; Conservative 17; Mismatches 41; Indels 30; Gaps
      1 GSLKSLSG---KGLQQRMFFLNDVDLLYTSRGLTASNQFVKVHQQLPLYGNTTIESEDEWG 57
      | | | | | : | | | | : | | | | : | | | |
      265 GWLLKLGSGRRVKTWKRWFILTDNCLYY----FEYTDKPRGIIIFLENLSTREVEDS-K 319
      | | | | | : | | | | : | | | | : | | | |
      58 VPHGLTL-----RGQRSTIIVAAASSRSEKWKVEDIOMA 91
      | | | | | : | | | | : | | | | : | | | |
      320 KPNCFELYIPDNKDQVIKACKTEADGRVVEGNHTVYRISAPTEERKEEWIKCIAA 375
      | | | | | : | | | | : | | | | : | | | |
  
```

ent NO. 6210914  
GENERAL INFORMATION:  
APPLICANT: Staunton, Donald E.  
APPLICANT: Lipsky, Brian P.  
TITLE OF INVENTION: Cytoplasmic Modulators of Integrin  
TITLE OF INVENTION: Binding/Signalling  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun  
STREET: 233 South Wacker, 6300 Sears Tower  
CITY: Chicago  
STATE: Illinois

ZIP: 00000  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/023.591A  
FILING DATE:

```
/
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/08/839,581
/ FILING DATE:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Young J. Suh
/ REGISTRATION NUMBER: P-41,337
/ REFERENCE/DOCKET NUMBER: 27866/33886
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 312-474-6300
/ TELEFAX: 312-474-0448
/ INFORMATION FOR SEQ ID NO: 32:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 398 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: single
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ US-09-023-591A-32

Query Match 15.2%; Score 71; DB 4; Length 398;
Best Local Similarity 24.1%; Pred. No. 0.24;
Matches 28; Conservative 17; Mismatches 41; Indels 30; Gaps 4;

QY 1 GSLSKLSG---KGLQRMFFLENDVLYTSRGLTASNQFKVHGQLPLYGMTIESEDEWG 57
| | | | | : | | | | | : | | | | | : | | | | |
Db 265 GWLLKLGGRVKTWKRWFILTDNCLY---FEYTTDKPRGIIPLENLSIREVEDS-K 319
| | | | | : | | | | | : | | | | | : | | | | |
QY 58 VPHCLTL-----RGQRQSIIVAASSRSEMEKWKVEDIQ 91
| | | | | : | | | | | : | | | | | : | | | | |
Db 320 KPNCFELYIPDNKDQVICKACKTEADGRVVGHNHTVYRISAPTEPEKEEWIKIAA 375
| | | | | : | | | | | : | | | | | : | | | | |

RESULT 13
US-09-284-033-5
; Sequence 5, Application US/09284033
; Patent No. 6194173
; GENERAL INFORMATION:
; APPLICANT: Czech, Michael P. and Klarlund, Jes K.
; TITLE OF INVENTION: BINDING PROTEINS FOR PHOSPHOINOSITIDES; GRPI OR GENERAL RECEP
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 28 STATE STREET
; CITY: BOSTON
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION NUMBER: US/09/284,033
; FILING DATE: 1999-04-06
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: USSN: 08/729,834
; FILING DATE: 07 OCTOBER 1996
; APPLICATION NUMBER: PCT/US97/18152
; FILING DATE: 1997-10-07
; ATTORNEY/AGENT INFORMATION:
; NAME: MANDRAGOURAS, AMY E.
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: UMM-018CPUS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)742-4214
; TELEFAX: (617)742-4214
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
```

```
/
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
/ FRAGMENT TYPE: internal
/ US-09-284-033-5

Query Match 15.1%; Score 70.5; DB 4; Length 136;
Best Local Similarity 24.8%; Pred. No. 0.059;
Matches 28; Conservative 16; Mismatches 40; Indels 29; Gaps 4;

QY 1 GSLSKLSG---KGLQRMFFLENDVLYTSRGLTASNQFKVHGQLPLYGMTIESEDEWG 58
| | | | | : | | | | | : | | | | | : | | | | |
Db 11 GWLLKLGGRVKTWKRWFILTDNCLY---FEYTTDKPRGIIPLENLSIREVEDP-RK 65
| | | | | : | | | | | : | | | | | : | | | | |
QY 59 PHCLTL-----RGQRQSIIVAASSRSEMEKWKVEDIQ 89
| | | | | : | | | | | : | | | | | : | | | | |
Db 66 PNCFELYNPESHKGQVICKACKTEADGRVVGHNHTVYRISAPSPKEKEEMKSIK 118
| | | | | : | | | | | : | | | | | : | | | | |

RESULT 14
US-08-729-834B-5
; Sequence 5, Application US/08729834B
; Patent No. 6221841
; GENERAL INFORMATION:
; APPLICANT: Czech, Michael P.
; APPLICANT: Klarlund, Jes K.
; TITLE OF INVENTION: General Receptors for Phosphoinositides
; TITLE OF INVENTION: and Uses Related Thereto
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 28 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION NUMBER: US/08/729,834B
; FILING DATE: October 7, 1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Amy E. Mandragouras
; REGISTRATION NUMBER: 36,207
; REFERENCE/DOCKET NUMBER: UMM-018
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)227-7400
; TELEFAX: (617)227-5941
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: Internal
; US-08-729-834B-5

Query Match 15.1%; Score 70.5; DB 4; Length 136;
Best Local Similarity 24.8%; Pred. No. 0.059;
Matches 28; Conservative 16; Mismatches 40; Indels 29; Gaps 4;

QY 1 GSLSKLSG---KGLQRMFFLENDVLYTSRGLTASNQFKVHGQLPLYGMTIESEDEWG 58
| | | | | : | | | | | : | | | | | : | | | | |
Db 11 GWLLKLGGRVKTWKRWFILTDNCLY---FEYTTDKPRGIIPLENLSIREVEDP-RK 65
| | | | | : | | | | | : | | | | | : | | | | |
QY 59 PHCLTL-----RGQRQSIIVAASSRSEMEKWKVEDIQ 89
| | | | | : | | | | | : | | | | | : | | | | |
```



Db 66 PNCFELNPSHKQVTKACKTEADGRVVGHNHVVYRISAPSPKEEWMKSIK 118

RESULT 15

US-09-284-033-2

; Sequence 2, Application US/09284033

; Patent No. 6194173

; GENERAL INFORMATION:

; APPLICANT: Czech, Michael P. and Klarlund, Jes K.

; TITLE OF INVENTION: BINDING PROTEINS FOR PHOSPHOINOSITIDES, GRP1 OR GENERAL RECEP

; NUMBER OF SEQUENCES: 17

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: LAHIVE & COCKFIELD, LLP

; STREET: 28 STATE STREET

; CITY: BOSTON

; STATE: MASSACHUSETTS

; COUNTRY: USA

; ZIP: 02109-1875

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/284,033

; FILING DATE: 1999-04-06

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: USSN: 08/729,834

; FILING DATE: 07 OCTOBER 1996

; APPLICATION NUMBER: PCT/US97/18152

; FILING DATE: 1997-10-07

; ATTORNEY/AGENT INFORMATION:

; NAME: MANDRAGOURAS, AMY E.

; REGISTRATION NUMBER: 36,207

; REFERENCE/DOCKET NUMBER: UMM-018CPUS

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (617)227-7400

; TELEFAX: (617)742-4214

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 399 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-284-033-2

Query Match 15.1%; Score 70.5; DB 4; Length 399;

Best Local Similarity 24.8%; Pred. No. 0.28;

Matches 28; Conservative 16; Mismatches 40; Indels 29; Gaps 4;

QY 1 GSLSKLSG--KGLQORMFLENDVLLYTSRGLTASNQPKVHGOLPLYGMTIESEDEWGV 58

DB 269 GWLLKLGGRVTKWRWFILTNCLY-----FEYTDKEPRGLIPLNLSIREVEDP-RK 323

QY 59 PHCLTL-----RGQRQSIIVAASSRSEMEKQWVEDIQ 89

DB 324 PNCFELNPSHKQVTKACKTEADGRVVGHNHVVYRISAPSPKEEWMKSIK 376

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:52:39 ; Search time 44.17 Seconds  
(without alignments)  
156.936 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_764\_854  
Perfect score: 468  
Sequence: 1 GSLSKLSGKGLQORFFLEFN.....IVAASSRSEMEKWKVEDIQMA 91  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68.\*  
1: pirl1.\*  
2: pirl2.\*  
3: pirl3.\*  
4: pirl4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	468	100.0	1045	2 JC5795	CDEP protein - hum
2	109.5	23.4	254	2 S40944	hypothetical prote
3	107	22.9	961	2 A55380	faciogenital dyspl
4	102	21.8	316	2 T13601	hypothetical prote
5	99.5	21.3	926	2 T15683	hypothetical prote
6	94.5	20.2	414	2 T22498	hypothetical prote
7	87	18.6	646	2 JC5583	85K SH3 domain-con
8	81	17.3	211	2 T17228	hypothetical prote
9	79.5	17.0	1244	2 S29083	guanine-nucleotide
10	79	16.9	994	1 JQ0151	myosin heavy chain
11	78.5	16.8	875	2 T19678	hypothetical prote
12	77	16.5	496	2 T46356	hypothetical prote
13	76.5	16.3	1275	2 A38985	nucleotide exchange
14	75	16.0	685	2 JC6331	rho-type guanine e
15	74	15.8	807	2 A34581	oxysterol-binding
16	74	15.8	809	2 A34404	oxysterol-binding
17	73.5	15.7	1557	2 T13160	protein CNK - frui
18	72.5	15.5	1260	2 S28407	guanine nucleotide
19	71	15.2	398	2 S24168	SEC7 homolog - hum
20	71	15.2	753	2 S48267	probable membrane
21	70.5	15.1	1737	2 A59235	unconventional myo
22	69.5	14.9	537	2 I49135	prolyl 4-hydroxyla
23	68	14.5	381	2 T46482	hypothetical prote
24	67.5	14.4	189	2 F82448	conserved hypothet
25	67.5	14.4	689	1 A53791	beta-adrenergic-re
26	66.5	14.2	380	2 C75016	na+/h+ antiporter
27	66.5	14.2	689	1 A40088	beta-adrenergic-re
28	66	14.1	252	2 E86104	ATP-binding compon
29	65.5	14.0	273	2 A86672	LysR family trans

30	65.5	14.0	409	2 D64216	hypothetical prote
31	65	13.9	648	2 T33339	hypothetical prote
32	64.5	13.8	374	2 JC7091	centaurin alpha 1
33	64.5	13.8	442	2 E71882	is606 transposase
34	64.5	13.8	570	2 T32743	hypothetical prote
35	64	13.7	230	2 T34541	hypothetical prote
36	63.5	13.6	488	2 T21596	hypothetical prote
37	63.5	13.6	566	2 T16375	son-of-sevenless-2
38	63	13.5	1297	2 S25714	hypothetical prote
39	62.5	13.4	359	2 F83850	hypothetical prote
40	62	13.2	168	2 T40346	ompl6 protein - Br
41	61.5	13.1	1166	2 S33812	myosin-like protei
42	61	13.0	275	2 F71516	probable diaminopi
43	61	13.0	513	2 S75997	asparagine-tRNA 1
44	61	13.0	611	1 A55888	protein kinase (EC
45	61	13.0	1115	2 T31068	N-methyl-D-asparta

ALIGNMENTS

RESULT 1  
JC5795  
CDEP protein - human  
C:Species: Homo sapiens (man)  
C:Date: 24-Jan-1998 #sequence\_revision 13-Mar-1998 #text\_change 16-Jul-1999  
C:Accession: JC5795  
R:Koyano, Y.; Kawamoto, T.; Shen, M.; Yan, W.; Noshiro, M.; Fujii, K.; Kato, Y.  
Biochem. Biophys. Res. Commun. 241, 369-375, 1997  
A:Title: Molecular cloning and characterization of CDEP, a novel human protein contain  
ing factors.  
A:Reference number: JC5795; MUID:98086358  
A:Accession: JC5795  
A:Molecule type: mRNA  
A:Residues: 1-1045 <KOY>  
A:Cross-references: DDBJ:AB008430  
C:Comment: This protein is involved in the adhesion, proliferation, and differentiati  
on.  
F:1-374/Domain: ezrin-like #status predicted <EZR>  
F:42-316/Domain: protein 4.1 membrane-binding domain homology <B41>  
F:931-1027/Domain: pleckstrin repeat homology <PLK>

Query Match	100.0%	Score	468	DB	2	Length	1045
Best Local Similarity	100.0%	Pred. No.	4.3e-45				
Matches	91	Conservative	0	Mismatches	0	Indels	0
Gaps	0						
Qy	1	GSLSKLSGKGLQORFFLEFN	DLVLLYTSGLTASNOFKVHGQLPLXGMTIESEDEWGVPH	60			
Db	764	GSLSKLSGKGLQORFFLEFN	DLVLLYTSGLTASNOFKVHGQLPLXGMTIESEDEWGVPH	823			
Qy	61	CLTLRGQRQSIIVAASSRSEMEKWKVEDIQMA	91				
Db	824	CLTLRGQRQSIIVAASSRSEMEKWKVEDIQMA	854				

RESULT 2  
S40944  
hypothetical protein ZK632.12 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 05-Dec-1998 #sequence\_revision 05-Dec-1998 #text\_change 05-Dec-1998  
C:Accession: S40944  
R:Berks, M.  
submitted to the EMBL Data Library, February 1993  
A:Reference number: S40933  
A:Accession: S40944  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-254 <BER>  
A:Cross-references: EMBL:Z22181  
C:Genetics:  
A:Introns: 6/1; 137/1; 206/3; 241/3



```

RESULT      7
JC5583
85K SH3 domain-containing proline-rich protein - mouse
C:Species: Mus musculus (house mouse)
C>Date: 23-Sep-1997 #sequence_revision 23-Sep-1997 #text_change 16-Dec-1998
C:Accession: JC5583
R:Oh, W.K.; Yoo, J.C.; Jo, D.; Song, Y.H.; Kim, M.G.; Park, D.
Biochem. Biophys. Res. Commun. 235, 794-798, 1997
A:Title: Cloning of a SH3 domain-containing proline-rich protein, p85SPR, and its localiza
A:Reference number: JC5583; MUID:97350865
A:Accession: JC5583
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-646 <OHA>
A:Cross-references: GB:U96634
A:Experimental source: thymus
C:Comment: This protein interacts with proteins in focal complexes.
C:Superfamily: SH3 homology; CDC24 homology
F:13-60/Domain: SH3 homology <SH3>
F:93-273/Domain: CDC24 homology <CD24>
F:F215-242,587-634/Domain: leucine zipper #status predicted <LZP>
F:276-279/Region: nuclear location signal
F:408-515/Region: proline-rich
F:508-511/Region: nuclear location signal

Query Match          18.6%; Score 87; DB 2; Length 646;
Best Local Similarity 28.0%; Pred. No. 0.065;
Matches 28; Conservative 21; Mismatches 37; Indels 14; Gaps 4;

Qy   1 GSLSKLSGKGGLQ-----QRMEFLF--NDVLLYTSRGILTASNQFKVHGQLPLYGMTIE 50
||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db   300 GSVTYMSQVTIQACGSEKNERYLLLPKPFSVMLSPSPRMXSG---FYQGKLPPTGTMIT 356

Qy   51 ESEDEGWVPCLTLRGOR-QSIIVAASSRSEMEEKWVEDIQ 89
: || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db   357 KLEDSENHRNAFEISGMIRILVSCYSQQDLHEWVEHLQ 396

RESULT      8
Tl7228
hypothetical protein DKFPz434D146.1 - human (fragment)
C:Species: Homo sapiens (man)
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: Tl7228
R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Cassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A:Reference number: ZlB722
A:Accession: Tl7228
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-211 <KOE>
A:Cross-references: EMBL:AL117429
A:Experimental source: adult testis; clone DKFPz434D146
C:Genetics:
A>Note: DKFPz434D146.1

```

```

RESULT      9
guanine-nucleotide-releasing protein - rat
S29083
N;Alternate names: CDC25 protein homolog
C;Species: Rattus norvegicus (Norway rat)
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 05-Nov-1999
C;Accession: S29083
R;Shou, C.; Farnsworth, C.L.; Neel, B.G.; Feig, L.A.
Nature 358, 351-354, 1992
A;Title: Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing factor fo
A;Reference number: S29083; MUID:923350260
A;Accession: S29083
A:Molecule type: mRNA
A;Residues: 1-1244 <SHO>
A;Cross-references: EMBL:X67241; NID:g57664; PIDN:CAA47666.1; PID:g57665
A;Note: The authors translated the codon GAG for residues 135 and 137 as Glu
C;Superfamily: CDC25-type guanine nucleotide exchange activator homology; CDC24 homol
F:240-426/Domain: CDC24 homology <CD24>
F:1005-1241/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match          17.0%; Score 79.5; DB 2; Length 1244;
Best Local Similarity 26.3%; Pred. No. 0.98;
Matches 29; Conservative 14; Mismatches 38; Indels 27; Gaps 3;

QY    1  GLSKLKSGKGLQRMFFLENDVLLYTSRGLTASNQFKVHGQLPLYGWTI---EESDEWG 57
       ||||| :||| |::| :||| :||| :||| :||| :||| :||| :||| :|||
Db    477  GRGLSLKKEGERQCFLFSKHLLICTRGSGSKLHLTKNGVISLIDCTLLDDPENMDDDG 536
                                           ::|
QY    58  VPCHLTLRGRQ-----SIIVAASSRSEMEKWVEDI 88
                                           :||
Db    537  -----KGQEVHLDPKIWPEKDSPFPTVLIVASSRQEKAWTSDI 577

RESULT     10
JQ0151
myosin heavy chain IA - slime mold (Dictyostellium sp.)
N;Alternate names: ambA protein
N;Contains: myosin ATPase (EC 3.6.1.32)
C;Species: Dictyostellium sp.
C;Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 02-Feb-2001
C;Accession: JQ0151
R;Titus, M.A.; Warrick, H.M.; Spudich, J.A.
Cell Regul. 1, 55-63, 1989
A;Title: Multiple actin-based motor genes in Dictyostellium.
A;Reference number: JQ0151; MUID:92096486
A;Accession: JQ0151
A:Molecule type: DNA
A;Residues: 1-994 <TIT>
A;Cross-references: GB:S73909; NID:g241267; PIDN:AAB20711.1; PID:g241268
A;Experimental source: strain A5-2
C;Genetics:
A;Gene: abma
A;Introns: 1/2; 42/3
C;Superfamily: brush border myosin heavy chain I; myosin motor domain homology
C;Keywords: actin binding; hydrolase; nucleotide binding; P-loop
F:15-706/Domain: myosin motor domain homology <MMOT>
F:105-112/Region: nucleotide-binding motif A (P-loop)
F:594-615/Region: actin binding #status predicted
F:783-994/Domain: carboxyl-terminal <CTD>
F:829-935/Region: basic
F:111/Binding site: ATP (Lys) #status predicted

Query Match          16.3%; Score 79; DB 1; Length 994;
Best Local Similarity 24.3%; Pred. No. 0.87;
Matches 25; Conservative 24; Mismatches 31; Indels 22; Gaps 5;

QY    3  LSKLSGKGLQRMFFLENDVLLYTSRG--LTASNQFKVHGQLPLYG--MTIESEDEWG 57
       ||::| |::| :|:| |::| |::| |::| |::| |::| |::| |::| :|
Db    20  LTVESSSLHENLKIRYEGLIYTSIGPVIVSMNPYK---OLGIYGNDQINLYKGHEFE 76
                                           :|
QY    58  VP-----HCLTLRGORQSIIVAASS---RSEMEXKV 85

```



RESULT 15  
 A34581  
 oxysterol-binding protein - human  
 C:Species: Homo sapiens (man)  
 C:Date: 22-Jun-1990 #sequence\_revision 22-Jun-1990 #text\_change 05-Nov-1999  
 C:Accession: A34581  
 R:Levanon, D.; Hsieh, C.L.; Francke, U.; Dawson, P.A.; Ridgway, N.D.; Brown, M.S.; Golds  
 Genomics 7, 65-74, 1990  
 A:Title: cDNA cloning of human oxysterol-binding protein and localization of the gene to  
 A:Reference number: A34581; MUID:90243258  
 A:Accession: A34581  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-807 <LEV>  
 A:Cross-references: GB:M86917; NID:g189402; PIDN:AAA59973.1; PID:g189403; GB:J04757  
 C:Genetics:  
 A:Gene: GDB:OSBP  
 A:Cross-references: GDB:120252; OMIM:167040  
 A:Map position: liql2-ilql3  
 C:Superfamily: pleckstrin repeat homology

Query Match 15.8%; Score 74; DB 2; Length 807;  
 Best Local Similarity 27.7%; Pred. No. 2.5;  
 Matches 26; Conservative 13; Mismatches 27; Indels 28; Gaps 4;

QY 9 KGLQRMFFLENDVLLY-----TSRGLTASNQFKVHGOLPLYGMTIESEDEWGV 59  
 DB 103 KGQRRWFVLSNGLLSYRSKAEHRHCTCGTI-----NLATANIIVEDS----- 146  
 QY 60 HC--LTLRGQRQSIIVAASSRSEMEKWEVDIQMA 91  
 DB 147 -CNFIISNGGAQTYHLKASSEVERQRWVTALELA 179

Search completed: December 6, 2001, 08:52:40  
 Job time: 372 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:53:10 ; Search time 28.69 Seconds  
(without alignments)  
116.295 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_764\_854

Perfect score: 468

Sequence: 1 GSLKSLGKGLQRMFLFN.....IVAASSRSEKWKVEDIQMA 91

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	109.5	23.4	266	1 YOTB_CAEEL	P34657 caenorhabdi
2	107	22.9	961	1 FGD1_HUMAN	P98174 homo sapien
3	106	22.6	960	1 FGD1_MOUSE	P52734 mus musculus
4	98	20.9	646	1 Y142_HUMAN	Q14155 homo sapien
5	84.5	18.1	394	1 CYH4_HUMAN	Q9ui40 homo sapien
6	79.5	17.0	1244	1 GNRP_RAT	P28818 rattus norv
7	79	16.9	994	1 MYSA_DICDI	P22467 dictyostell
8	76.5	16.3	1275	1 GNRP_HUMAN	Q13972 homo sapien
9	75	16.0	400	1 CYH2_HUMAN	Q99418 homo sapien
10	75	16.0	400	1 CYH2_MOUSE	P97695 mus musculus
11	74	15.8	807	1 OXYB_HUMAN	P22059 homo sapien
12	74	15.8	809	1 OXYB_RABIT	P16258 oryctolagus
13	72.5	15.5	1262	1 GNRP_MOUSE	P27671 mus musculus
14	71	15.2	398	1 CYH1_HUMAN	Q15438 homo sapien
15	71	15.2	753	1 YB2_YEAST	P38261 saccharomyc
16	70.5	15.1	399	1 CYH3_MOUSE	O08967 mus musculus
17	70	15.0	400	1 CYH3_HUMAN	O43739 homo sapien
18	70	15.0	400	1 CYH3_RAT	P97696 rattus norv
19	69.5	14.9	537	1 P4H2_MOUSE	Q60716 mus musculus
20	69	14.7	398	1 CYH1_MOUSE	Q9qx11 mus musculus
21	69	14.7	398	1 CYH1_RAT	P97694 rattus norv
22	67.5	14.4	689	1 ARK1_HUMAN	P25098 homo sapien
23	67	14.3	638	1 Y053_HUMAN	P42331 homo sapien
24	66.5	14.2	689	1 ARK1_BOVIN	P21146 bos taurus
25	66.5	14.2	689	1 ARK1_MESAU	Q64682 mesocricetu
26	65.5	14.0	409	1 Y148_MYCCE	P47394 mycoplasma
27	64	13.7	1332	1 SOS2_HUMAN	Q07890 homo sapien
28	63	13.5	1297	1 SOS2_MOUSE	Q02384 mus musculus
29	61	13.0	275	1 DAPF_CHLTR	O84437 chlamydia t
30	61	13.0	513	1 SYN_SYNY3	P52276 synechocyst
31	60.5	12.9	207	1 NEF_HV1ND	P18801 human immun
32	60.5	12.9	397	1 PEL_TOBAC	P40972 nicotiana t
33	60.5	12.9	860	1 ENV_HV2BE	P18094 human immun

34 60.5 12.9 1045 1 MYHL\_CHICK P47807 gallus gall  
35 60.5 12.9 1107 1 MYIA\_MOUSE P46735 mus musculus  
36 60.5 12.9 1136 1 MYIA\_RAT Q05096 rattus norv  
37 60 12.8 252 1 PHNK\_ECOLI P16678 escherichia  
38 60 12.8 496 1 LA12\_ARATH Q06402 arabidopsis  
39 60 12.8 626 1 GCKR\_RAT Q07071 rattus norv  
40 59.5 12.7 206 1 NEF\_HV1EL P04604 human immun  
41 59.5 12.7 480 1 S61A\_YEAST P32915 saccharomyc  
42 59.5 12.7 689 1 ARK1\_RAT P28617 rattus norv  
43 59 12.6 505 1 MALQ\_STRPN P29851 streptococc  
44 58.5 12.5 519 1 TIM\_HUMAN Q12774 homo sapien  
45 58.5 12.5 688 1 ARK2\_RAT P26819 rattus norv

## ALIGNMENTS

RESULT 1  
YOTB\_CAEEL STANDARD; PRT; 266 AA.  
AC P34657;  
DT 01-FEB-1994 (Rel. 28, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE HYPOTHETICAL 30.2 KDA PROTEIN ZK632.12 IN CHROMOSOME III.  
GN ZK632.12.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton S., Du Z., Durbin R., Copsey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Durrin R., Favello A., Fraser A.,  
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,  
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkneen R.,  
RA Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,  
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
RA Waterson R., Watson A., Weinstein L., Wilkinson-Sproat J.,  
RA Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.";  
RL Nature 368:32-38(1994).  
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 FYVE-TYPE ZINC FINGER.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; Z22181; CA880187.1;  
DR WormPep; ZK632.12; CE01110.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR000306; znf\_FYVE.  
DR Pfam; PF01363; FYVE; 1.  
DR Pfam; PF00169; PH; 1.  
DR SMART; SM00064; FYVE; 1.  
DR SMART; SM00233; PH; 1.  
DR PROSITE; PS50178; ZF\_FYVE; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
KW Hypothetical protein; Zinc-finger.  
FT DOMAIN 35 131  
FT ZN\_FING 152 212 FYVE-TYPE.  
SQ SEQUENCE 266 AA; 30187 MW; 91C2F62EDF13839E CRC64;

```
DR PROSITE; PS50003; PH_DOMAIN; 2.
KW Guanine-nucleotide releasing factor; Zinc-finger.
FT DOMAIN 7 330 PRO-RICH.
FT DOMAIN 171 179 SH3-BINDING (POTENTIAL).
FT DOMAIN 179 187 SH3-BINDING (POTENTIAL).
FT DOMAIN 590 689 PH 1.
FT ZN_FING 730 790 FYVE-TYPE.
FT DOMAIN 821 921 PH 2.
SQ SEQUENCE 961 AA; 106668 MW; CF722598853A685A CRC64;

Query Match 22.9%; Score 107; DB 1; Length 961;
Best Local Similarity 33.0%; Pred. No. 0.00018;
Matches 30; Conservative 17; Mismatches 40; Indels 4; Gaps 3;

QY 1 GSLSKLSGKG--LQRMFFLENDVLLYTSRGLTASNQ-FKHVHQLPLYGTMITESEDEWG 57
DB 595 GHILKLSAKNGTTQDRYLILFNRLLYCVPRLLRGKFSVRARIDVDGMELKESN-LN 653

QY 58 VPHCLTLRGQRQSIIVAASRSSEMEKWVEDI 88
DB 654 LPPTFLVSGKQSRSLQARTTEEEKDWQVAI 684

RESULT 3
FGDL_MOUSE
ID FGDL_MOUSE STANDARD; PRT; 960 AA.
AC P52734;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)
DE (FACIOGENITAL DYSPLASIA PROTEIN HOMOLOG).
GN FGDL_MOUSE
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96081343; PubMed=8535076;
RA Pasteris N.G., de Gouyon B., Cadle A.B., Campbell K., Herman G.E.,
RA Gorski J.L.;
RT "Cloning and regional localization of the mouse facio-genital
RT dysplasia (Fgdl) gene.";
RL Mamm. Genome 6:658-661(1995).
CC -!- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS
CC BY EXCHANGING BOUND GDP FOR FREE GTP.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 FYVE-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U11690; AAA57004.1; -
CC MIM; 305400; -
CC InterPro; IPR001849; PH.
CC InterPro; IPR000219; RhoGEF.
CC InterPro; IPR000822; Znf-C2H2.
CC InterPro; IPR000306; Znf_FYVE.
CC Pfam; PF01363; FYVE; 1.
CC Pfam; PF00169; PH; 2.
CC Pfam; PF00621; RhoGEF; 1.
CC SMART; SM00233; PH; 2.
CC SMART; SM00325; RhoGEF; 1.
CC PROSITE; PS50178; ZF_FYVE; 1.

Query Match 23.4%; Score 109.5; DB 1; Length 266;
Best Local Similarity 30.3%; Pred. No. 2e-05;
Matches 27; Conservative 21; Mismatches 40; Indels 1; Gaps 1;

QY 1 GSLSKLSGKGLQRMFFLENDVLLYTSRGLTASNQFKVHQLPLYGTMITESEDEWGVPH 60
DB 40 GVLVWCKRKPKQRFLENDILVY-GNIVISKRYNKQRILEGVQVEDLEDGIGKH 98

QY 61 CLTLRGQRQSIIVAASRSSEMEKWVEDIQ 89
DB 99 GWIITPAKSFAYAAATEKREWMHLIE 127

RESULT 2
FGDL_HUMAN
ID FGDL_HUMAN STANDARD; PRT; 961 AA.
AC P98174;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)
DE (FACIOGENITAL DYSPLASIA PROTEIN).
GN FGDL_HUMAN
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95042764; PubMed=7954831;
RA Pasteris N.G., Cadle A., Logie L.J., Porteous M.E.M., Schwartz C.E.,
RA Stevenson R.E., Glover T.W., Wilroy R.S., Gorski J.L.;
RT "Isolation and characterization of the facio-genital dysplasia
RT (Aarskog-Scott syndrome) gene: a putative Rho/Rac guanine nucleotide
RT exchange factor.";
RL Cell 79:669-678(1994).
CC -!- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS
CC BY EXCHANGING BOUND GDP FOR FREE GTP.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: FETAL HEART, BRAIN, LUNG, KIDNEY, PLACENTA,
CC AND LESS IN LIVER; ADULT HEART, BRAIN, LUNG, SKELETAL MUSCLE, AND
CC LESS IN PANCREAS AND LIVER.
CC -!- DISEASE: FACIOGENITAL DYSPLASIA (AARSKOG-SCOTT SYNDROME) IS A RARE
CC MULTISYSTEMIC DISORDER CHARACTERIZED BY DISPROPORTIONATELY SHORT
CC STATURE, AND BY FACIAL, SKELETAL, AND UROGENITAL ANOMALIES.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 FYVE-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U11690; AAA57004.1; -
CC MIM; 305400; -
CC InterPro; IPR001849; PH.
CC InterPro; IPR000219; RhoGEF.
CC InterPro; IPR000822; Znf-C2H2.
CC InterPro; IPR000306; Znf_FYVE.
CC Pfam; PF01363; FYVE; 1.
CC Pfam; PF00169; PH; 2.
CC Pfam; PF00621; RhoGEF; 1.
CC SMART; SM00233; PH; 2.
CC SMART; SM00325; RhoGEF; 1.
CC PROSITE; PS50178; ZF_FYVE; 1.
```

```
DR PROSITE; PS50178; 2F_FYVE; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
KW Guanine-nucleotide releasing factor; Zinc-finger.
FT DOMAIN 7 330 PRO-RICH.
FT DOMAIN 171 179 SH3-BINDING (POTENTIAL).
FT DOMAIN 179 187 SH3-BINDING (POTENTIAL).
FT DOMAIN 589 688 PH 1.
FT ZN_FING 729 789 FYVE-TYPE.
FT DOMAIN 820 920 PH 2.
SQ SEQUENCE 960 AA; 106477 MW; 41CLB84DE49FC51 CRC64;

Query Match 22.6%; Score 106; DB 1; Length 960;
Best Local Similarity 33.0%; Pred. No. 0.00023;
Matches 30; Conservative 17; Mismatches 40; Indels 4; Gaps 3;

QY 1 GSLSKLSGKG--LQORMFLEFNDVLLYTSRGLTASNQ-FKVGQPLPLGYMTIESEDEWG 57
Db 594 GHTLKUSANGTTQDRYLILFNDRLLYCVPRLLGOKFTVRARIDVDGMELKESN-LN 652

QY 58 VPHCLRLGRQSIIVAAASSRSEMEKRWEDI 88
Db 653 MPRTFLVSGKQSRLEQLQARTEEKDWQAI 683

RESULT 4
Y142_HUMAN
ID Y142_HUMAN STANDARD; PRT; 646 AA.
AC Q14155;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0142.
GN KIAA0142.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96127530; PubMed=8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res 2:167-174(1995).
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63476; BAA09763.1; -.
DR HSSP; P19174; LHSQ.
DR InterPro; IPR001849; PH.
DR InterPro; IPR002119; RhGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Hypothetical protein; SH3 domain.
FT DOMAIN 6 65 SH3.
```

```
FT DOMAIN 295 400 PH.
SQ SEQUENCE 646 AA; 73140 MW; B5B5A83F0EBC28D2 CRC64;

Query Match 20.9%; Score 98; DB 1; Length 646;
Best Local Similarity 32.1%; Pred. No. 0.0012;
Matches 25; Conservative 17; Mismatches 34; Indels 2; Gaps 2;

QY 13 QRMFLEFNDVLLYTSRGLTASNQFKVHGQPLPLGYMTIESEDEWGPVHCLTLRQGR-QSI 71
Db 320 ERYLLFLFPNVLLMLSLASPRMSG-FIYQGLPTTGMITKLEDSHNHNAFEISGSMIERI 378

QY 72 IVAASSRSEMEKRWEDIQ 89
Db 379 LVSCNQDQLQEWVHLQ 396

RESULT 5
CYH4_HUMAN
ID CYH4_HUMAN STANDARD; PRT; 394 AA.
AC Q9UIA0; Q9UGT6;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOHESIN 4.
GN PSCD4 OR CYT4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RA Liu D., Zhang H., Lu J.;
RT "cDNA cloning and genomic organization of cytohesin-4";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20119275; PubMed=10652308;
RA Ogasawara M., Kim S.C., Adamik R., Togawa A., Ferrans V.J., Takeda K.,
RA Kirby M., Moss J., Vaughan M.;
RT "Similarities in function and gene structure of cytohesin-4 and
RT cytohesin-1, guanine nucleotide-exchange proteins for ADP-ribosylation
RT factors.";
RL J. Biol. Chem. 275:3221-3230(2000).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA Clamp M., Smink L.J., Alnsough R., Almeida J.P., Babbage A.,
RA Bagguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
RA Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
RA Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
RA Coville G.J., Cox A.V., Davis J., Dawson E., Dhani P.D., Dockree C.,
RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
RA Fleming K French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
RA Grafham D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,
RA Heathcote R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
RA Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,
RA Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
RA Matthews L., McCann O.T., Mcclay J., McLaren S., Mcmuray A.A.,
RA Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
RA Pearson D., Phillimore B.J., Phillips S.H., Plumb R.W., Ramsay H.,
RA Ramsey Y., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
RA Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
RA Sulston J.E., Swann R.M., Vaubin M., Wall M., Wallis J.M.,
RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
RA Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
RA Do A., Do T., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
```

RA Lai H., Lao H.I., Lewis J., Lewis S., Lin S.-P., Loh P., Mala J. E.,  
RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,  
RA Sloyan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,  
RA Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,  
RA Minx P., Fulton R., Johnson D., Benis G., Bentley D., Bradshaw H.,  
RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,  
RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozerisky P.,  
RA Rohlfing T., Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K.,  
RA Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,  
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,  
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,  
RA Edelmann L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,  
RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,  
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,  
RA Khan A.S., Lane L., Tilahun Y., Wright H.,  
RT "The DNA sequence of human chromosome 22.";  
RL Nature 402:489-495(1999).  
CC -!- FUNCTION: PROMOTES GUANINE-NUCLEOTIDE EXCHANGE ON ARF1 AND ARF5.  
CC PROMOTES THE ACTIVATION OF ARF THROUGH REPLACEMENT OF GDP WITH  
CC GTP.  
CC -!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN PERIPHERAL BLOOD  
CC LEUKOCYTES.  
CC -!- SIMILARITY: CONTAINS 1 SEC7 DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL: AF075458; AAF15389.1; -  
DR EMBL: AF123349; AAF28896.1; -  
DR EMBL: Z94160; GAB3067.1; -  
DR HSSP: Q99418; IPBV.  
DR InterPro: IPR001849; PH.  
DR InterPro: IPR000904; Sec7.  
DR Pfam: PF00169; PH; 1.  
DR Pfam: PF01369; Sec7; 1.  
DR SMART: SM00233; PH; 1.  
DR SMART: SM00222; Sec7; 1.  
DR PROSITE: PS50003; PH\_DOMAIN; 1.  
DR PROSITE: PS50190; SEC7; 1.  
KW Guanine-nucleotide releasing factor; Coiled coil.  
FT DOMAIN 12 65 COILED COIL (POTENTIAL).  
FT DOMAIN 72 201 SEC7.  
FT DOMAIN 259 375 PH.  
SQ SEQUENCE 394 AA; 45672 MW; 0C9C8E9AD6890F34 CRC64;  
  
Query Match 18.1%; Score 84.5; DB 1; Length 394;  
Best Local Similarity 25.7%; Pred. No. 0.025;  
Matches 29; Conservative 22; Mismatches 33; Indels 29; Gaps 5;  
  
QY 1 GSLSKLSG--KGLQRMFFLNDVLLYTSRGLTASNQFKVHGOLPLYGMTIESEDEWG 58  
Db 264 GLLKLGGRVKTWKRWFILTDNCLY-----FEFTDKPRGLIPLNLSVQKVDPP-KK 318  
QY 59 PHCLTL-----RQQR-----QSIIVAASSRSEMEKWKVEDIQ 89  
Db 319 PFCLELYNPSCRQKTKACKTDGDRVVEGKESYRISATSABERDQWIESIR 371  
  
RESULT 6  
GNRP\_RAT STANDARD; PRT; 1244 AA.  
AC P28818;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE GUANINE NUCLEOTIDE RELEASING PROTEIN (GNRP) (P140 RAS-GRF).

GN RASGRF1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=92350260; PubMed=1379346;  
RA Shou C., Farusworth C.L., Neel B.G., Feig L.A.;  
RT "Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing  
RL factor for Ras p21.";  
RL Nature 358:351-354(1992).  
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.  
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.  
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL: X67241; CAA47666.1; -  
DR PIR: S29083; S29083; GDS\_CDC24.  
DR InterPro: IPR001331; GDS\_CDC24.  
DR InterPro: IPR000048; IQ.  
DR InterPro: IPR001849; PH.  
DR InterPro: IPR000651; RasGEFN.  
DR InterPro: IPR001895; RasGEF\_CDC25.  
DR InterPro: IPR000219; RHOGEF.  
DR Pfam: PF00612; IQ; 1.  
DR Pfam: PF00169; PH; 2.  
DR Pfam: PF00617; RasGEF; 1.  
DR Pfam: PF00618; RasGEFN; 1.  
DR Pfam: PF00621; RHOGEF; 1.  
DR SMART: SM00015; IQ; 1.  
DR SMART: SM00233; PH; 2.  
DR SMART: SM00147; RasGEF; 1.  
DR SMART: SM00229; RasGEFN; 2.  
DR SMART: SM00325; RHOGEF; 1.  
DR PROSITE: PS00720; GDS\_CDC25; 1.  
DR PROSITE: PS00741; GDS\_CDC24; 1.  
DR PROSITE: PS50003; PH\_DOMAIN; 2.  
DR PROSITE: PS50096; IQ; 1.  
KW Guanine-nucleotide releasing factor.  
FT DOMAIN 22 129  
FT DOMAIN 204 229 IQ.  
FT DOMAIN 244 455 DH.  
FT DOMAIN 456 582 PH 2.  
FT DOMAIN 1007 1241 RASGEF.  
SQ SEQUENCE 1244 AA; 142666 MW; 4B647879E842AF6B CRC64;  
  
Query Match 17.0%; Score 79.5; DB 1; Length 1244;  
Best Local Similarity 26.9%; Pred. No. 0.37;  
Matches 29; Conservative 14; Mismatches 38; Indels 27; Gaps 3;  
  
QY 1 GSLSKLSGKGLQRMFFLNDVLLYTSRGLTASNQFKVHGOLPLYGMTI---EESDEWG 57  
Db 477 GRGLSLKLKKEGRCQFLFSKHLIICTRGSGSKLHLTKNGVISLIDCTLDLDDFNWDDG 536  
QY 58 VPHCLTLRQQR-----SIIVAASSRSEMEKWKVEDI 88  
Db 537 -----KQGEVDHDLFKIWVEPKDPPPTVILVASSRQEKAAWTSDI 577  
  
RESULT 7  
MYSALDICI

```

ID MYSA_DICDI STANDARD; PRT; 994 AA.
AC P22467;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MYOSIN IA HEAVY CHAIN (MYOSIN-LIKE PROTEIN ABMA).
GN ABMA OR DMTA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A5-2;
RX MEDLINE=92096486; PubMed=2519618;
RA Titus M.A., Warrick H.M., Spudich J.A.;
RT "Multiple actin-based motor genes in Dictyostelium.";
RL Cell Regul. 1:55-63(1989).
CC -!- FUNCTION: ACTIN-BASED MOTOR PROTEIN, POSSIBLY INVOLVED IN A WIDE
CC RANGE OF MOTILE PROCESSES, SUCH AS CELL MOVEMENT ACROSS A SURFACE,
CC AND EXTENSION AND RETRACTION OF PSEUDOPODIA OR LAMELLIPODIA.
CC -!- SUBUNIT: MYOSIN I HEAVY CHAIN IS SINGLE-HEADED. DIMER OF A HEAVY
CC AND A LIGHT CHAIN. INABILITY TO SELF-ASSEMBLE INTO FILAMENTS.
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 2 IQ DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; S73909; AAB20711.1; -.
CC PIR; JQ0151; JQ0151.
CC HSP; P08799; 1MND.
CC DictyDb; DD01028; abma.
CC InterPro; IPR000048; IQ.
CC InterPro; IPR001609; myosin_head.
CC Pfam; PF00612; IQ; 2.
CC PRINTS; PR00193; MYOSINHEAVY.
CC ProDom; PD000355; myosin_head; 1.
CC SMART; SM00013; IQ; 1.
CC SMART; SM00242; MYSC; 1.
CC PROSITE; PS50096; IQ; 1.
CC Myosin; Actin-binding; ATP-binding; Calmodulin-binding;
CC Multigene family.
CC DOMAIN 1 743 MYOSIN HEAD-LIKE.
CC DOMAIN 744 773 IQ.
CC DOMAIN 774 994 NON ALPHA-HELICAL, C-TERMINAL DOMAIN.
CC NP_BIND 105 112 ATP (POTENTIAL).
CC DOMAIN 575 655 ACTIN-BINDING.
CC DOMAIN 830 931 MEMBRANE-BINDING (POTENTIAL).
CC SEQUENCE 994 AA; 113286 MW; D30A5D20885B118C CRC64;
SQ
Query Match 16.9%; Score 79; DB 1; Length 994;
Best Local Similarity 24.5%; Pred. No. 0.32;
Matches 25; Conservative 24; Mismatches 31; Indels 22; Gaps 5;
QY 3 LSKLSKGLQORFFLNDVLLYTSRG--LTASNQKFKVHQQLPLYG---MTIESEDEWG 57
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 20 LTVESSSLHENLKIRYKEGLIYTSIGPLVSMNPYK---QLGIYNDQINLYKGRHEFE 76
QY 58 VP-----HCLTLRGQRSIIIVASS-----RSEMEKW 85
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 77 IPPIIYIADKAYRALRSEGENOCIIISGSGAGKTEASKYI 118
RESULT 8
GNRP_HUMAN STANDARD; PRT; 1275 AA.
ID GNRP_HUMAN
```

```

AC Q13972;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GUANINE NUCLEOTIDE RELEASING PROTEIN (GNRP) (RAS-SPECIFIC NUCLEOTIDE
DE EXCHANGE FACTOR CDC25).
GN RASGRF1 OR CDC25.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wei W., Broek D.;
RT "Cloning and analysis of the full length human cdc25 cDNA, a ras-
RT specific nucleotide exchange factor.";
RL Submitted (DEC-1993) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; L26584; AAA58417.1; -.
CC InterPro; IPR001331; GDS_CDC24.
CC InterPro; IPR000048; IQ.
CC InterPro; IPR001849; PH.
CC InterPro; IPR000651; RasGEFN.
CC InterPro; IPR001895; RasGRF_CDC25.
CC InterPro; IPR000219; RhoGEF.
CC Pfam; PF00612; IQ; 1.
CC Pfam; PF00617; RasGEF; 1.
CC Pfam; PF00618; RasGEFN; 1.
CC Pfam; PF00621; RhoGEF; 1.
CC SMART; SM00015; IQ; 1.
CC SMART; SM00233; PH; 2.
CC SMART; SM00147; RasGEF; 1.
CC SMART; SM00229; RasGEFN; 2.
CC SMART; SM00325; RhoGEF; 1.
CC PROSITE; PS00720; GDS_CDC25; 1.
CC PROSITE; PS00741; GDS_CDC24; 1.
CC PROSITE; PS50003; PH_DOMAIN; 2.
CC PROSITE; PS50096; IQ; 1.
CC Guanine-nucleotide releasing factor.
CC DOMAIN 22 129 PH 1.
CC DOMAIN 204 229 IQ.
CC DOMAIN 247 431 DH.
CC DOMAIN 467 584 PH 2.
CC DOMAIN 1038 1272 RASGEF.
CC SEQUENCE 1275 AA; 145381 MW; 86C6F54AA1E451F1 CRC64;
SQ
Query Match 16.3%; Score 76.5; DB 1; Length 1275;
Best Local Similarity 25.2%; Pred. No. 0.85;
Matches 28; Conservative 14; Mismatches 38; Indels 31; Gaps 3;
QY 1 GSKLSKGLQORFFLNDVLLYTSRGLTASNQKFKVHQQLPLYGNTI-----ESEDEW 56
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 477 GRGLSLEKEGERQCFLFKSHLIICITRGSGGKLHLTKNGVISLIDCTLLEEPSTEE- 535
QY 57 GVPHCLTLRGQRO-----SIIVAASRSSEMEKWVEDI 88
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 536 -----AKSGQDIDHDFKIGVEPKDSPFPFTVLVASSRQKAAWTSDI 579
```

```
RESULT 9
CYH2_HUMAN STANDARD; PRT; 400 AA.
AC Q99418; Q92958;
DT 15-JUL-1998 (Rel. 36, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOHESIN 2 (ARF NUCLEOTIDE-BINDING SITE OPENER) (ARNO PROTEIN) (ARF
DE EXCHANGE FACTOR).
GN PSCD2 OR ARNO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RA Chardin P., Paris S., Antony B., Robineau S., Bernaud-Dufour S.,
RA Jackson C.L., Chabre M.;
RT "A human exchange factor for ARF contains Sec7 and pleckstrin-homology
RT domains.";
RL Nature 384:481-484(1996).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 2), AND CHARACTERIZATION.
RC TISSUE=Brain;
RX MEDLINE=98079021; PubMed=9417041;
RA Frank S.F., Upender S.K., Hansen S.H., Casanova J.E.;
RT "ARNO is a guanine nucleotide exchange factor for ADP-ribosylation
RT factor 6.";
RL J. Biol. Chem. 273:23-27(1998).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS) OF 51-252.
RX MEDLINE=98135767; PubMed=9476900;
RA Mossessova E., Gulbis J.M., Goldberg J.;
RT "Structure of the guanine nucleotide exchange factor Sec7 domain of
RT human ARNO and analysis of the interaction with ARF GTPase.";
RL Cell 92:415-423(1998).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 51-252.
RX MEDLINE=98169075; PubMed=9510236;
RA Cherifis J., Menetrey J., Mathieu M., le Bras G., Robineau S.,
RA Beraud-Dufour S., Antony B., Chardin P.;
RT "Structure of the Sec7 domain of the Arf exchange factor ARNO.";
RL Nature 392:101-105(1998).
CC -!- FUNCTION: PROMOTES GUANINE-NUCLEOTIDE EXCHANGE ON ARF1, ARF3 AND
CC ARF6. PROMOTES THE ACTIVATION OF ARF THROUGH REPLACEMENT OF GDP
CC WITH GTP.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: CONTAINS 1 SEC7 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X99753; CAA68084.1; -
DR EMBL; U70728; AAB09591.1; -
DR PDB; 1PBV; 09-MAR-99.
DR MIM; 602488; -
DR InterPro; IPR001849; PH.
DR InterPro; IPR000904; Sec7.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF01369; Sec7; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00222; Sec7; 1.
```

```
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50190; SEC7; 1.
KW Guanine-nucleotide releasing factor; Coiled coil;
KW Alternative splicing; 3D-structure.
FT DOMAIN 10 63 COILED COIL (POTENTIAL).
FT DOMAIN 72 201 SEC7.
FT DOMAIN 259 376 PH.
FT VARSPLIC 272 272 MISSING (IN ISOFORM 2).
SQ SEQUENCE 400 AA; 46546 MW; 70441A58483BD0E1 CRC64;

Query Match 16.0%; Score 75; DB 1; Length 400;
Best Local Similarity 24.3%; Pred. No. 0.32;
Matches 28; Conservative 21; Mismatches 38; Indels 28; Gaps 5;

QY 1 GSLSKLSG---KGLQRMFLFNDVLLYTSRGLTASNOFKVHQLPLYGNTIEESD--- 54
DB 264 GLLKLGGRVKTWKRWFILTONCLYY-----FEYTTDKREPRGIILENISIREVDDPRK 319
QY 55 ----EWGVPH-----CLT-----LRGQRQSIIVAAASSRSEMEKWKVEDIQMA 91
DB 320 PNCFELYIPNNKQLIKACKTEADGRVVEGNHMYRISAPTQEKDEWIKSIQAA 374

RESULT 10
CYH2_MOUSE STANDARD; PRT; 400 AA.
AC P97695; O89099;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOHESIN 2 (ARF NUCLEOTIDE-BINDING SITE OPENER) (ARNO PROTEIN) (CLM2)
DE (SEC7 HOMOLOG B) (MSEC7-2).
GN PSCD2 OR SEC7B.
OS Mus musculus (Mouse), and
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090, 10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).
RC SPECIES=Mouse; TISSUE=Brain;
RX MEDLINE=98416066; PubMed=9744817;
RA Kim H.-S., Chen Y., Lonai P.;
RT "Complex regulation of multiple cytohesin-like genes in murine tissues
RT and cells.";
RL FEBS Lett. 433:312-316(1998).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC SPECIES=Mouse;
RA Liu D., Zhang H., Lu J.;
RT "cDNA cloning of mouse cytohesin-2 and demonstration of its
RT association with the integrin beta2 subunit.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC SPECIES=Rat;
RX MEDLINE=98013473; PubMed=9352219;
RA Telmerakis I., Benseler F., Stenius K., Sudhof T.C., Brose N.;
RT "Rat homologues of yeast sec7p.";
RL Eur. J. Cell Biol. 74:143-149(1997).
CC -!- FUNCTION: PROMOTES GUANINE-NUCLEOTIDE EXCHANGE ON ARF1, ARF3 AND
CC ARF6. PROMOTES THE ACTIVATION OF ARF THROUGH REPLACEMENT OF GDP
CC WITH GTP.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1/CLM2-A (SHOWN HERE), 2 AND
CC 3/CLM2-B; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: PRESENT IN ALL TISSUES TESTED, WITH HIGHEST
CC PROTEIN LEVELS IN BRAIN AND ADRENAL.
CC -!- SIMILARITY: CONTAINS 1 SEC7 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```



```
CC EMBL; J05056; AAA31427.1; -.
DR PIR; A34404.
DR InterPro; IPR000648; Oxysterol_BP.
DR InterPro; IPR001849; PH.
DR Pfam; PF01237; Oxysterol_BP; 1.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00233; PH; 1.
DR PROSITE; PS01013; OSBP; 1.
DR MOD_RES; 1 1 BLOCKED.
FT DOMAIN 1 92 GLY/ALA-RICH.
FT DOMAIN 90 183 PH.
SQ SEQUENCE 809 AA; 89478 MW; 55C6CAE1B985B1E0 CRC64;

Query Match 15.8%; Score 74; DB 1; Length 809;
Best Local Similarity 27.7%; Pred. No. 0.96;
Matches 26; Conservative 13; Mismatches 27; Indels 28; Gaps 4;

QY 9 KGLQRMFFELNDVLLY-----TSRGLTASNQFKVHGQLPLYGMTIESEDEWGV 59
Db 105 KGQRRWFVLSNGLLSYRKAEMRHTCGTI-----NLATANTIVDS----- 148

QY 60 HC--LTLRGQRQSIIVASRSSEMEKWKVEDIQMA 91
Db 149 -CNFIISNGGAQTVHLKASSEVERQRWVTALELA 181

RESULT 13
ID GNRP_MOUSE STANDARD; PRT; 1262 AA.
AC P27671;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE GUANINE NUCLEOTIDE RELEASING PROTEIN (GNRP) (RAS-SPECIFIC NUCLEOTIDE
DE EXCHANGE FACTOR CDC25) (CDC25MM).
GN RASGRF1 OR CDC25 OR GRF1.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC SEQUENCE FROM N.A.
CP STRAIN-BALB/C.
RX MEDLINE=93010996; PubMed=1396590;
RA Cen H., Lowy D.D.;
RT "Isolation of multiple mouse cDNAs with coding homology to
RT Saccharomyces cerevisiae CDC25: identification of a region related to
RT Bcr, Vav, Dbl and CDC24."
RL EMBO J. 11:4007-4015(1992).
RN [2]
RP SEQUENCE OF 791-1262 FROM N.A.
RC STRAIN=SWISS; TISSUE=Brain;
RX MEDLINE=92289680; PubMed=1376246;
RA Martegani E., Vanoni M., Zippel R., Coccetti P., Brambilla R.,
RA Ferrari C., Sturani E.P., Alberghina L.;
RT "Cloning by functional complementation of a mouse cDNA encoding a
RT homologue of CDC25, a Saccharomyces cerevisiae RAS activator."
RL EMBO J. 11:2151-2157(1992).
RN [3]
RP SEQUENCE OF 1031-1226 FROM N.A.
RX MEDLINE=92357779; PubMed=1379731;
RA Wei W., Mosceller R.D., Sanyal P., Gonzales E., McKinney D.,
RA Dasgupta C., Li P., Liu B.X., Broek D.;
RT "Identification of a mammalian gene structurally and functionally
RT related to the CDC25 gene of Saccharomyces cerevisiae."
RL Proc. Natl. Acad. Sci. U.S.A. 89:7100-7104(1992).
CC -1- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
```

```
CC -1- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L20899; AAA02741.1; -.
DR EMBL; X59868; CAA42525.1; -.
DR PIR; S20730; S20730.
DR PIR; S22693; S22693.
DR MGD; MGI:99694; Rasgrfl.
DR InterPro; IPR001331; GDS_CDC24.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000651; RasGEFN.
DR InterPro; IPR001895; RasGEF_CDC25.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00169; PH; 2.
DR Pfam; PF00617; RasGEF; 1.
DR Pfam; PF00618; RasGEFN; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEFN; 2.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00720; GDS_CDC25; 1.
DR PROSITE; PS00741; GDS_CDC24; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
DR PROSITE; PS50096; IQ; 1.
KW Guanine-nucleotide releasing factor.
FT DOMAIN 22 130 PH 1.
FT DOMAIN 208 233 IQ.
FT DOMAIN 248 459 DH.
FT DOMAIN 460 588 PH 2.
FT DOMAIN 1025 1259 RASGEF.
FT CONFLICT 1033 1033 E -> D (IN REF. 3).
SQ SEQUENCE 1262 AA; 144101 MW; 38BFE68F7C228DC8 CRC64;

Query Match 15.5%; Score 72.5; DB 1; Length 1262;
Best Local Similarity 27.2%; Pred. No. 2.4;
Matches 28; Conservative 13; Mismatches 47; Indels 15; Gaps 3;

QY 1 GSLSKLSGKGLQRMFFELNDVLLYTSRGLTASNQFKVHGQLPLYGMTI---EESSEW 56
Db 481 GRLGSLTKKEGGRQCFLFSKHLIICTRGSGGKHLTKNGVISLIDCTLDEPENLDEA 540

QY 57 -----GVPHCLTLRGQRQ-----SIIVASRSSEMEKWKVEDI 88
Db 541 KGAGPETEHLFEKIGVPEKDSLPTFTVLVASTROEKAAWTSDI 583

RESULT 14
CYH1_HUMAN
ID CYH1_HUMAN STANDARD; PRT; 398 AA.
AC Q15438; O9P123; O9P124;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE CYTOHESIN 1 (SEC7 HOMOLOG B2-1).
GN PCSDI OR D17S811E.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:54:35 ; Search time 77.11 Seconds  
(without alignments)  
172.621 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_764\_854

Perfect score: 468

Sequence: 1 GSIKSLGKGLQRMFLFN.....IVAASSRSEKWKVEDIQMA 91

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.17.\*  
1: sp-archaea.\*  
2: sp-bacteria.\*  
3: sp-fungi.\*  
4: sp-human.\*  
5: sp-invertebrate.\*  
6: sp-mammal.\*  
7: sp-mhc.\*  
8: sp-organalle.\*  
9: sp-phage.\*  
10: sp-plant.\*  
11: sp-rodent.\*  
12: sp-virus.\*  
13: sp-vertebrate.\*  
14: sp-unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	468	100.0	1045	Q9Y4F1	Q9Y4f1 homo sapien
2	461	98.5	420	Q99K03	Q99ku3 mus musculus
3	287	61.3	1054	Q94887	Q94887 homo sapien
4	152.5	32.6	432	Q9NUR8	Q9nur8 homo sapien
5	152.5	32.6	699	Q9P215	Q9p215 homo sapien
6	149.5	31.9	431	Q9D3Y7	Q9d3y7 mus musculus
7	122	26.1	766	Q88387	Q88387 rattus norv
8	119	25.4	621	Q93504	Q93504 brachydanio
9	118.5	25.3	279	Q95UB9	Q95ub9 homo sapien
10	116.5	24.9	279	Q99M16	Q99m16 mus musculus
11	112	23.9	727	Q88841	Q88841 mus musculus
12	111.5	23.8	249	Q9H8W4	Q9h8w4 homo sapien
13	111.5	22.8	562	Q9VN69	Q9vn69 drosophila
14	107	22.9	397	Q99K79	Q99k79 mus musculus
15	104	22.2	1161	Q9NEH0	Q9neh0 leishmania
16	102	21.8	316	Q76902	Q76902 drosophila
17	99.5	21.3	926	Q18284	Q18284 caenorhabdi
18	99	21.2	698	Q9NXY1	Q9nxy1 homo sapien
19	96	20.5	646	Q55043	Q55043 rattus norv

Query Match 100.0%; Score 468; DB 4; Length 1045;  
Best Local Similarity 100.0%; Pred. No. 1.5e-46;

20	94.5	20.2	414	5	Q20653	Q20653 caenorhabdi
21	88	18.8	1028	4	Q9UGQ4	Q9ugq4 homo sapien
22	88	18.8	1539	4	Q9P2D2	Q9p2d2 homo sapien
23	87	18.6	630	11	Q9ES27	Q9es27 mus musculus
24	87	18.6	646	11	O08757	O08757 mus musculus
25	87	18.6	705	11	Q9ES28	Q9es28 mus musculus
26	86	18.4	733	11	O88842	O88842 mus musculus
27	85.5	18.3	762	4	Q9H961	Q9h961 homo sapien
28	84.5	18.1	740	4	Q9UPP2	Q9upp2 homo sapien
29	81.5	17.4	201	4	Q9NUL6	Q9nul6 homo sapien
30	81	17.3	211	4	Q9UFW5	Q9ufw5 homo sapien
31	81	17.3	301	4	Q9H9R2	Q9h9r2 homo sapien
32	81	17.3	408	4	Q9H9D7	Q9h9d7 homo sapien
33	80.5	17.2	554	11	Q9JHT9	Q9jht9 mus musculus
34	80	17.1	755	11	Q9CXJ3	Q9cxj3 mus musculus
35	79	16.9	1355	4	Q9HCK9	Q9hck9 homo sapien
36	78.5	16.8	875	5	Q18372	Q18372 caenorhabdi
37	78	16.7	400	11	Q9EQD2	Q9equ2 apodemus ag
38	77	16.5	496	4	Q9NTG0	Q9ntg0 homo sapien
39	77	16.5	619	4	Q9HDC6	Q9hdc6 homo sapien
40	77	16.5	694	4	Q9UPP0	Q9upp0 homo sapien
41	77	16.5	720	4	Q9NR80	Q9nr80 homo sapien
42	77	16.5	1210	5	Q9VP80	Q9vp80 drosophila
43	75.5	16.1	399	11	Q99KH2	Q99kh2 mus musculus
44	75.5	16.1	1194	4	Q9P2F6	Q9p2f6 homo sapien
45	75	16.0	266	4	Q15795	Q15795 homo sapien

## ALIGNMENTS

RESULT 1

Q9Y4F1 PRELIMINARY; PRT; 1045 AA.  
 AC Q9Y4F1;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)  
 DE CDEP;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-CARTILAGE;  
 RX MEDLINE=98086358; PubMed=9425278;  
 RA Koyano Y., Kawamoto T., Shen M., Yan W., Noshiro M., Fujii K.,  
 RA Kato Y.;  
 RT "Molecular cloning and characterization of CDEP, a novel human protein  
 RT containing the ezrin-like domain of the band 4.1 superfamily and the  
 RT Dbl homology domain of Rho guanine nucleotide exchange factors.";  
 RL Biochem. Biophys. Res. Commun. 241:369-375(1997).  
 DR EMBL; AB008430; BAA24267.1;  
 DR HSSP; P08567; IPLS.  
 DR InterPro; IPR000299; Band\_4.1.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR000219; RhoGEF.  
 DR Pfam; PF00373; Band\_41; 1.  
 DR Pfam; PF00169; PH; 2.  
 DR Pfam; PF00621; RhoGEF; 1.  
 DR PRINTS; P00935; BAND41.  
 DR SMART; SM00295; B41; 1.  
 DR SMART; SM00233; PH; 2.  
 DR SMART; SM00325; RhoGEF; 1.  
 DR PROSITE; PS00560; BAND\_41\_1; UNKNOWN\_1.  
 DR PROSITE; PS00557; BAND\_41\_3; 1.  
 DR PROSITE; PS00503; PH\_DOMAIN; 2.  
 SQ SEQUENCE 1045 AA; 118632 MW; 0E8B2D61C0F58417 CRC64;





```
O88387
ID O88387 PRELIMINARY; PRT; 766 AA.
AC O88387
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ACTIN-FILAMENT BINDING PROTEIN FRABIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98334590; PubMed=9668039;
RA Obalishi H., Nakanishi H., Mandai K., Satoh K., Satoh A., Takahashi K.,
RA Miyahara M., Nishiohara H., Takai Y.,
RT "Frabin, a novel FGD1-related actin filament-binding protein capable
RT of changing cell shape and activating c-Jun N-terminal kinase."
RL J. Biol. Chem. 273:18697-18700(1998).
DR EMBL: AF038388; AAC27698.1; --
DR SMART: SM00233; PH; 2.
DR INTERPRO: IPR000219; RHOGEF.
DR INTERPRO: IPR000306; Znf_FYVE.
DR INTERPRO: IPR000822; Znf_C2H2.
DR INTERPRO: IPR001849; PH.
DR PFAM: PF00169; PH; 2.
DR PFAM: PF00621; RHOGEF; 1.
DR PFAM: PF01363; FYVE; 1.
DR PROSITE: PS00003; PH_DOMAIN; 2.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
DR SMART: SM00233; PH; 2.
DR SMART: SM00325; RHOGEF; 1.
DR SMART: SM00325; RHOGEF; 1.
SQ SEQUENCE 766 AA; 86453 MW; E34831DC0B0B334C CRC64;

Query Match 26.1%; Score 122; DB 11; Length 766;
Best Local Similarity 31.5%; Pred. No. 7.4e-06;
Matches 29; Conservative 24; Mismatches 35; Indels 4; Gaps 3;

QY 1 GSLSKLSGK--GLQORMFFLNDVLLY--TSRGLTASNQFKVHGOLPLYGMTIESEDEWG 57
Db 427 GQILKLAARNTSAOERLYLFNNMLYCVPRFLVSGKFTVTRVRGIDGKIVETHNE-E 485

QY 58 VPCHLTILRGQRQSIIVAAASRSEMEKVEDIQ 89
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : : ||
Db 486 YPHTFVSGKERTLEQASSEQKKEWIKALQ 517

QY 58 VPCHLTILRGQRQSIIVAAASRSEMEKVEDIQ 89
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : : ||
Db 486 YPHTFVSGKERTLEQASSEQKKEWIKALQ 517

RESULT 8
O93504 PRELIMINARY; PRT; 621 AA.
AC O93504
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE FACIOGENITAL DYSPLASIA PROTEIN.
GN FGD.
OS Brachydanio rerio (Zebrafish) (Zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Rasbora; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Pasteris N.G., Buckler J., Gorski J.L.;
RT "Isolation and Characterization of a zebrafish homologue of the
RT Faciogenital Dysplasia (FGD1) Gene."
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF017370; AAC35432.1; --
DR INTERPRO: IPR000219; RHOGEF.
DR INTERPRO: IPR000306; Znf_FYVE.
DR INTERPRO: IPR001849; PH.
DR PFAM: PF00169; PH; 1.
```

```
DR Pfam: PF00621; RHOGEF; 1.
DR Pfam: PF01363; FYVE; 1.
DR PROSITE: PS00003; PH_DOMAIN; 1.
DR SMART: SM00064; FYVE; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RHOGEF; 1.
SQ SEQUENCE 621 AA; 70852 MW; BA5B35BD1B2A08BB CRC64;

Query Match 25.4%; Score 119; DB 13; Length 621;
Best Local Similarity 30.8%; Pred. No. 1.3e-05;
Matches 28; Conservative 22; Mismatches 37; Indels 4; Gaps 3;

QY 1 GSLSKLSGK--GLQORMFFLNDVLLY--TSRGLTASNQFKVHGOLPLYGMTIESEDEWG 57
Db 380 GHKKWSAKNGSAQDRYLYLFNNMLYCVPKRLMGOKFSVREIRIDTAGMEVHNVKQ-N 438

QY 58 VPCHLTILRGQRQSIIVAAASRSEMEKVEDIQ 88
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : : ||
Db 439 VPHTFTITGKRSLELQARTAEERDDWIKVI 469

RESULT 9
O9BUB9 PRELIMINARY; PRT; 279 AA.
ID O9BUB9
AC O9BUB9
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:4090).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=NEUROBLASTOMA;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC003744; AAH02744.1; --
SQ SEQUENCE 279 AA; 31194 MW; ED2ADEBDE6BCF3BF CRC64;

Query Match 25.3%; Score 118.5; DB 4; Length 279;
Best Local Similarity 29.9%; Pred. No. 5.7e-06;
Matches 29; Conservative 25; Mismatches 26; Indels 17; Gaps 4;

QY 1 GSLSKLSGKGLQORMFFLNDVLLY--TSRGLTASNQFKVHGQ--LPLYGMTIE-----ES 52
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : : ||
Db 40 GVLTKECRKKAKPRIFLFDILVY---GSIVLNKKRYSQHIIPLEVTLELLPETLQA 96

QY 53 EDEWGVPHCLTLRGQRQSIIVAAASRSEMEKVEDIQ 89
|| : : : : : || : : : : : || : : : : : || : : : : : || : : : : : ||
Db 97 KNRW-----MIKTAKKSFVVSAAATERQEWISHIE 127

RESULT 10
O99M16 PRELIMINARY; PRT; 279 AA.
ID O99M16
AC O99M16
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:6627).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC002120; AAH02120.1; --
```

```
SQ SEQUENCE 279 AA; 31200 MW; 14F3B2B9AAF96DD5 CRC64;

Query Match      24.9%; Score 116.5; DB 11; Length 279;
Best Local Similarity 28.4%; Pred. No. 9.9e-06;
Matches 27; Conservative 28; Mismatches 27; Indels 13; Gaps 3;

QY 1 GSLSKSLGKGLQORWFFLENDVLLYTSRGLTASNQKFKVHGQPLVGMTE-----ESED 54
DB 40 GVLTRECKRKAKPRIFFLNDLVGSIIVL-SKKRYRSQHIIPLFVTLPLPETLQAKN 98
QY 55 ENGVPCHLTQRQOSIIVAASSRSEMEKWKVEDIQ 89
DB 99 RW-----MIKTAKKSFVVSAASTTERQEWISHIE 127

RESULT 11
ID O88841 PRELIMINARY; PRT; 727 AA.
AC O88841.
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DE 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE FACIOGENITAL DYSPLASIA PROTEIN 2.
GN FGD2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Pasteris N.G., Gorski J.L.;
RT "Isolation, characterization and mapping of Fgd2, a homologue of the
RT facio-genital dysplasia (FGD1; Aarskog syndrome) gene.";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF017368; AAC35430.1; -.
DR MGD: MGI:1347084; Fgd2.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR000306; Znf_FYVE.
DR Pfam: PF01363; FYVE; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhoGEF; 1.
DR SMART: SM00064; FYVE; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhoGEF; 1.
DR PROSITE: PS00003; PH_DOMAIN; 1.
SQ SEQUENCE 727 AA; 82099 MW; 5CD0FBFB905C5FC5 CRC64;

Query Match      23.9%; Score 112; DB 11; Length 727;
Best Local Similarity 33.8%; Pred. No. 0.00011;
Matches 27; Conservative 17; Mismatches 34; Indels 2; Gaps 2;

QY 13 QRMFFLENDVLLY-TSRGLTASNQKFKVHGQPLVGMTEESDEGVPCHLTQRQSI 71
DB 338 ERYLVFNMLLYCVPRVLQVGAQVQVTRIDVAGMKVRELTD-AEPHSFLVSGKQRTL 396
QY 72 IVAASSRSEMEKWKVEDIQ 91
DB 397 ELQARSDEMSWMAQCA 416

RESULT 12
ID Q9H8W4 PRELIMINARY; PRT; 249 AA.
AC Q9H8W4.
DT 01-MAR-2001 (TremBLrel. 16, Created)
DT 01-MAR-2001 (TremBLrel. 16, Last sequence update)
DE 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE CDNA FLJ13187 FIS, CLONE NT2RP3004242, WEAKLY SIMILAR TO PUTATIVE
DE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR.
OS Homo sapiens (Human).
```

```
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isoqai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoliri T., Raku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saïto K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwavanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK023249; BAB14486.1; -.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000306; Znf_FYVE.
DR Pfam: PF01363; FYVE; 1.
DR Pfam: PF00169; PH; 1.
DR SMART: SM00064; FYVE; 1.
DR SMART: SM00233; PH; 1.
DR PROSITE: PS00003; PH_DOMAIN; 1.
SQ SEQUENCE 249 AA; 27797 MW; F5E3F84595A98886 CRC64;

Query Match      23.8%; Score 111.5; DB 4; Length 249;
Best Local Similarity 31.8%; Pred. No. 3.4e-05;
Matches 28; Conservative 20; Mismatches 39; Indels 1; Gaps 1;

QY 1 GSLSKSLGKGLQORWFFLENDVLLYTSRGLTASNQKFKVHGQPLVGMTEESDEGVPCH 60
DB 40 GVLTCLCRKKRPFLENDLVGNIQKKYKQH-IIPLENTIDSIKDEGLRN 98
QY 61 CLTLRQROSIIVAAASSRSEMEKWKVEDI 88
DB 99 GWLIKTPTKSFVAYATATEKSEMNNHI 126

RESULT 13
ID Q9VN69 PRELIMINARY; PRT; 562 AA.
AC Q9VN69.
DT 01-MAY-2000 (TremBLrel. 13, Created)
DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
DE 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE CG2008 PROTEIN.
GN CG2008.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abri J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
```





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:51:42 ; Search time 72.4 Seconds  
(without alignments)  
198.484 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_544\_737

Perfect score: 1023

Sequence: 1 IAEVSTERTYKLDLEIVT.....AEITEMVAQLHGTMRKMFN 194

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A\_Geneseq\_1101.\*  
1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1023	100.0	1045	20 AAY07482	Human chondrocyte-
2	1023	100.0	1045	21 AAY91947	Human cytoskeleton
3	924	90.3	395	21 AAB54227	Human pancreatic c
4	195.5	19.1	766	21 AAY51248	Rat actin-binding
5	176	17.2	523	22 AAM39338	Human polypeptide
6	176	17.2	647	22 AAM41124	Human polypeptide
7	171	16.7	619	22 AAB97025	Human colon carcin
8	169	16.5	1683	21 AAY71160	Rat phosphodiester
9	168.5	16.5	1658	21 AAY57450	Mouse Ese2L protei
10	163	15.9	1715	21 AAY57449	Mouse Esi1L protei
11	160.5	15.7	844	21 AAB36516	Candida albicans C

12	149	14.6	1520	20 AAY41010	Amino acid sequenc
13	147	14.4	1527	22 AAU01184	Rat glutamate tran
14	138	13.5	2861	18 AAW27227	Human TRIO phospho
15	131	12.8	1461	19 AAW64468	Human secreted pro
16	131	12.8	1461	22 AAB90743	Human CW420.2 pro
17	125	12.2	846	22 AAM41561	Human polypeptide
18	122.5	12.0	1237	22 AAM39825	Human polypeptide
19	122.5	12.0	1257	22 AAM41611	Human polypeptide
20	122	11.9	580	20 AAW81349	Human guanine nucl
21	120	11.7	1244	21 AAY68825	Amino acid sequenc
22	118.5	11.6	798	22 AAB94519	Human protein seq0
23	118	11.5	477	15 AAB66067	Phosphorylated p66
24	117.5	11.5	1227	22 AAB65637	Novel protein kina
25	117.5	11.5	1289	20 AAY27163	Peptide Seq ID No:
26	117.5	11.5	1289	21 AAY56781	Human Trad protein
27	113.5	11.1	1269	22 AAM40297	Human polypeptide
28	113.5	11.1	1315	22 AAM42083	Human polypeptide
29	110	10.8	550	20 AAW81351	Human guanine nucl
30	110	10.8	1604	22 AAB47327	FCR4. Homo sapie
31	108	10.6	2596	22 AAB30569	A splice variant o
32	106	10.4	845	21 AAY49419	PKA substrate, Vav
33	105	10.3	827	22 AAM40424	Human polypeptide
34	105	10.3	827	22 AAM40425	Human polypeptide
35	105	10.3	872	20 AAY07073	Renal cancer assoc
36	104	10.2	391	21 AAB37413	Human secreted pro
37	103.5	10.1	847	20 AAY22237	Human KDR signal t
38	102	10.0	408	22 AAB95198	Human protein sequ
39	98	9.6	378	21 AAB54156	Human pancreatic c
40	98	9.6	596	21 AAB01203	Human GTPase assoc
41	97	9.5	235	20 AAW81350	Human guanine nucl
42	97	9.5	327	20 AAW73380	CSB2 protein sequ
43	97	9.5	384	21 AAB37414	Human secreted pro
44	97	9.5	412	22 AAB95879	Human protein sequ
45	97	9.5	526	20 AAW73378	CSB2 protein sequ

#### ALIGNMENTS

RESULT	1
AAV07482	
ID	AAV07482 standard; Protein; 1045 AA.
XX	AAV07482;
AC	AAV07482;
DT	17-AUG-1999 (first entry)
XX	Human chondrocyte-derived protein CDEP.
DE	
XX	
KW	Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;
KW	Dbl homology domain; plectstrin homology domain; rheumatoid 'arthritis;
KW	drug.
XX	
OS	Homo sapiens.
XX	
PN	WO9928458-Al.
XX	
PD	10-JUN-1999.
XX	
PF	27-NOV-1998; 98WO-JP05348.
XX	
PR	27-NOV-1997; 97JP-0342060.
XX	(CHUS ) CHUGAI SEIYAKU KK.
PA	
XX	
PI	Kato Y, Kawamoto T, Koyano Y;
XX	
DR	WPI; 1999-371117/31.
XX	N-PSDB; AAX79183.
PT	Protein CDEP expressed in differentiated chondrocytes, and gene
PT	encoding it
XX	

PS Claim 2; Fig 1; 59pp; Japanese.

XX This sequence represents a protein (CDEP) expressed in differentiated

CC human foetal chondrocytes, which contains an ezrin-like domain, a Dbl

CC homology (DH) domain and a pleckstrin homology (PH) domain. The encoding

CC nucleic acid or protein can be used in the investigation and treatment of

CC cancers and arthritic diseases (including chronic rheumatoid arthritis),

CC or for screening of candidate anticancer drugs.

XX

SQ Sequence 1045 AA;

Query Match 100.0%; Score 1023; DB 20; Length 1045;

Best Local Similarity 100.0%; Pred. NO. 1.4e-98;

Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 IAEVSTERTYIKDLEVTISWQSVSKEDAMPEALKSLIFNPFELKHFHTNFIKEIE 60

Db 544 IAEVSTERTYIKDLEVTISWQSVSKEDAMPEALKSLIFNPFELKHFHTNFIKEIE 603

QY 61 QRLALWEGRSNAQIRDYQRIGDVMKNGKHLAAHLWKHSEALENGIKSSRRLEN 120

Db 604 QRLALWEGRSNAQIRDYQRIGDVMKNGKHLAAHLWKHSEALENGIKSSRRLEN 663

QY 121 FCRDFELQKVCYILPLNTFLRLPLRLMHYKQVLERLCKHHPPSHADFRDCRAALAEITEM 180

Db 664 FCRDFELQKVCYILPLNTFLRLPLRLMHYKQVLERLCKHHPPSHADFRDCRAALAEITEM 723

QY 181 VAQLHGTMIKMF 194

Db 724 VAQLHGTMIKMF 737

RESULT 2

AA91947

ID AAY91947 standard; Protein; 1045 AA.

XX

AC AAY91947;

XX

DT 19-JUL-2000 (first entry)

XX

DE Human cytoskeleton associated protein 2 (CYSKP-2).

XX

KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;

KW autoimmune; inflammatory, vesicle trafficking; neurological;

KW cardiovascular; cell motility; reproductive; muscle disorder.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT 23

FT Modified-site /note= "potential phosphorylation site"

FT 36

FT Modified-site /note= "potential phosphorylation site"

FT 41

FT Modified-site /note= "potential phosphorylation site"

FT Domain 47..85

FT /note= "signature sequence"

FT 92

FT Modified-site /note= "potential phosphorylation site"

FT 94..123

FT Domain /note= "signature sequence"

FT 144..190

FT Domain /note= "signature sequence"

FT 150

FT Modified-site /note= "potential phosphorylation site"

FT 152

FT Modified-site /note= "potential N-glycosylation site"

FT Domain 196..249

FT /note= "signature sequence"

FT 207

FT Modified-site /note= "potential phosphorylation site"

FT 261..279

FT Domain /note= "potential phosphorylation site"

FT 270

FT Modified-site /note= "signature sequence"

FT 336

FT Modified-site /note= "potential phosphorylation site"

FT 340

FT Modified-site /note= "potential phosphorylation site"

FT 343

FT Modified-site /note= "potential phosphorylation site"

FT 366

FT Modified-site /note= "potential phosphorylation site"

FT 370

FT Modified-site /note= "potential phosphorylation site"

FT 396

FT Modified-site /note= "potential phosphorylation site"

FT 408

FT Modified-site /note= "potential phosphorylation site"

FT 418

FT Modified-site /note= "potential phosphorylation site"

FT 448

FT Modified-site /note= "potential phosphorylation site"

FT 495

FT Modified-site /note= "potential N-glycosylation site"

FT 525

FT Modified-site /note= "potential phosphorylation site"

FT 538

FT Modified-site /note= "potential phosphorylation site"

FT 542

FT Modified-site /note= "potential phosphorylation site"

FT 549

FT Modified-site /note= "potential phosphorylation site"

FT 551

FT Modified-site /note= "potential phosphorylation site"

FT 571

FT Modified-site /note= "potential phosphorylation site"

FT 657

FT Modified-site /note= "potential phosphorylation site"

FT 658

FT Modified-site /note= "potential phosphorylation site"

FT 706

FT Modified-site /note= "potential phosphorylation site"

FT 770

FT Modified-site /note= "potential phosphorylation site"

FT 789

FT Modified-site /note= "potential phosphorylation site"

FT 811

FT Modified-site /note= "potential phosphorylation site"

FT 815

FT Modified-site /note= "potential phosphorylation site"

FT 826

FT Modified-site /note= "potential phosphorylation site"

FT 839

FT Modified-site /note= "potential phosphorylation site"

FT 770..773

FT Domain /note= "signature sequence"

FT 840

FT Modified-site /note= "potential phosphorylation site"

FT 842

FT Modified-site /note= "potential phosphorylation site"

FT 872

FT Modified-site /note= "potential phosphorylation site"

FT 878

FT Modified-site /note= "potential phosphorylation site"

FT 883

FT Modified-site /note= "potential phosphorylation site"

FT 889

FT Modified-site /note= "potential phosphorylation site"

FT 898

FT Modified-site /note= "potential phosphorylation site"

FT 919

FT Modified-site /note= "potential N-glycosylation site"

FT 923

FT Modified-site /note= "potential phosphorylation site"

FT	Modified-site	966	/note= "potential phosphorylation site"
XX	WO200017355-A2.		
XX	30-MAR-2000.		
XX	17-SEP-1999;	99WO-US21565.	
XX	18-SEP-1998;	98US-0172226.	
XX	27-APR-1999;	99US-0131321.	
XX	(INCY-) INCYTE PHARM INC.		
XX	Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;		
PI	Guegler KJ, Patterson C, Azimzai Y, Baughn MR;		
XX	WPI: 2000-283582/24.		
DR	N-PSDB; AAA08582.		
XX	Human cytoskeleton associated proteins, used to treat cell		
PT	proliferative, autoimmune/inflammatory, vesicle trafficking,		
PT	neurological, cell motility, reproductive and muscle disorders		
XX	Claim 1; Page 82-84; 113pp; English.		
XX	AAV91946-61 show human cytoskeleton associated proteins 1 to 16 (CYSKP-1		
CC	to CYSKP-16) respectively. The sequences can be used to treat and		
CC	diagnose cancer and cell proliferative, autoimmune/inflammatory, vesicle		
CC	trafficking, neurological, cardiovascular, cell motility, reproductive		
CC	and muscle disorders. Pharmaceutical compositions containing CYSKP-2 to		
CC	CYSKP-16 can be used to treat or prevent disorders associated with		
CC	decreased expression or activity of CYSKP (claimed), for example,		
CC	atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,		
CC	cancers, autoimmune/antiflammatory disorders such as allergies, anemia,		
CC	asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,		
CC	diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma		
CC	and trauma. CYSKP antagonists can be used to treat or prevent a		
CC	disorder associated with increased expression or activity of CYSKP		
XX	(claimed).		
XX	Sequence 1045 AA;		
SQ			
Query Match 100.0%; Score 1023; DB 21; Length 1045;			
Best Local Similarity 100.0%; Pred. No. 1.4e-98;			
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	IAKEVSTTERTYKDLVITSNFQSTVSKEDAMPEALKSLIFPNFPLHKFHTNFKETE	60
Db	544	IAKEVSTTERTYKDLVITSNFQSTVSKEDAMPEALKSLIFPNFPLHKFHTNFKETE	603
QY	61	QLALWEGRSNAQIRYQRTGDVLMKNIQGMKHLAHLWKHSEALENGIKSSRRLN	120
Db	604	qlalwegrnagirdyqrigdvmlknigmkhlahlwkhsalealengikssrllen	663
QY	121	FCRDFELQKVCYVLPNFTLRLPLHRLMHYKQVLERLCKHPPSHADFCRAALAEITEM	180
Db	664	fcrdfelqkvclplntfllrplrhlmhykqvlrclckhppshadfcraalaeitem	723
QY	181	VAQLHGTMIKMFNF 194	
Db	724	vaqlhgtmikmf 737	
RESULT 3			
ID	AA54227	standard; Protein; 395 AA.	
XX	AC		
XX	AA54227;		
DT	09-MAR-2001	(first entry)	
XX			
DE	Human pancreatic cancer antigen protein sequence SEQ ID NO:679.		
XX	Human; pancreas; pancreatic cancer; pancreatic cancer antigen;		
KW	detection; diagnosis; identification; cytostatic; neuroprotective;		
KW	nootropic; immunomodulatory; relaxant; contraceptive; gynaecological;		
KW	antiinflammatory; cardiant; gene therapy; chromosome mapping;		
KW	linkage analysis; tissue identification; tissue typing; forensic;		
KW	neural; immune system; muscular; reproductive; gastrointestinal;		
KW	pulmonary; cardiovascular; renal; proliferative.		
OS	Homo sapiens.		
XX	WO200055320-A1.		
PN	21-SEP-2000.		
PD	08-MAR-2000; 2000WO-US05989.		
PF	12-MAR-1999; 99US-0124270.		
PR	(HUMA-) HUMAN GENOME SCI INC.		
XX	Rosen CA, Ruben SM;		
PI	WPI: 2000-579444/54.		
DR	N-PSDB; AAC98992.		
XX	New nucleic acid that is a pancreatic cancer antigen for preventing,		
PT	treating, or ameliorating a medical condition, particular pancreatic		
PT	cancer, or for use in assays for diagnosing a pathological condition -		
XX	Claim 11; Page 1115-1116; 1379pp; English.		
XX	AAC98773 to AAC99231 encode the human pancreatic cancer associated		
CC	proteins, called pancreatic cancer antigens, given in AAB54008 to		
CC	AAB54466. The human pancreatic cancer antigens have cytostatic,		
CC	neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive,		
CC	gynaecological, cardiant and antiinflammatory activities, and can be used		
CC	in gene therapy. The polynucleotide and proteins can be used for		
CC	preventing, treating, or ameliorating a medical condition or in assays		
CC	for diagnosing a pathological condition or a susceptibility to one in a		
CC	subject. Binding partners to the proteins and the activity of the		
CC	proteins can be identified. The pancreatic cancer antigens can be used to		
CC	detect, treat or prevent pancreatic disorders, especially cancer.		
CC	Agonists and antagonists to the antigens can be screened for. The		
CC	pancreatic cancer antigen polynucleotides can be used to design nucleic		
CC	acid hybridisation probes that can be used in chromosome mapping, linkage		
CC	analysis, tissue identification and/or typing and a variety of forensic		
CC	and diagnostic methods. The proteins can be used to generate antibodies		
CC	which are used to purify, detect and target the polypeptides, including		
CC	both in vivo and in vitro diagnostic and therapeutic methods. The		
CC	proteins can be used to treat or prevent neural, immune system, muscular,		
CC	reproductive, gastrointestinal, pulmonary, cardiovascular, renal or		
CC	proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent		
CC	sequences used in the exemplification of the present invention.		
XX	Sequence 395 AA;		
SQ			
Query Match 90.3%; Score 924; DB 21; Length 395;			
Best Local Similarity 100.0%; Pred. No. 9.6e-89;			
Matches 173; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	22	WFQSTVSKEDAMPEALKSLIFPNFPLHKFHTNFKETEQLALWEGRSNAQIRYQIG	81
Db	21	wfqstvsksedampealkslifpnfplhkftnflkeieqlalwegrnagirdyqrig	80
QY	82	DVNLKNIQGMKHLAHLWKHSEALENGIKSSRRLNFCRDFELQKVCYVLPNFTLRL	141
Db	81	dvmlknigmkhlahlwkhsalealengikssrrlncrdfelqkvclplntfllr	140
QY	142	PLHRLMHYKQVLERLCKHPPSHADFCRAALAEITEMVAQLHGTMIKMFNF 194	
XX			

Db 141 plhlrhykqvlrclckhppsbadfrdcraaelaitemvqqlhgtmknemf 193

## RESULT 4

AA151248  
ID AAY51248 standard; Protein; 766 AA.

XX  
AC AAY51248;

XX  
DT 14-APR-2000 (first entry)

XX  
DE Rat actin-binding protein frabin.

XX  
KW Actin-binding protein; frabin; rat; developmental stage; animal growth.

XX  
OS Rattus sp.

XX  
PN JP11346775-A.

XX  
PD 21-DEC-1999.

XX  
PF 11-JUN-1998; 98JP-0164016.

XX  
PR 11-JUN-1998; 98JP-0164016.

XX  
PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX  
PA (OBAT/) OBAISHI H.

XX  
DR WPI; 2000-109691/10.

XX  
DR N-PSDB; AA244678.

XX  
PT An actin-binding protein frabin - plays an important role in the

XX  
PS developmental stages of animal growth

XX  
PS Claim 1; Page 6-8; 12pp; Japanese.

XX  
CC This invention describes a novel actin-binding protein, frabin (I). The

XX  
CC new actin-binding protein frabin plays an important role in the

XX  
CC developmental stages of animal growth. The protein can be used in gene

XX  
CC engineering. This sequence represents the rat frabin protein described

XX  
CC in the method of the invention.

XX  
SQ Sequence 766 AA;

Query Match 19.1%; Score 195.5; DB 21; Length 766;

Best Local Similarity 26.3%; Pred. No. 8.4e-12; Mismatches 90; Indels 13; Gaps 4;

Matches 52; Conservative 43; Mismatches 90; Indels 13; Gaps 4;

QY 1 IAKEVSTTERTYKDLVITSWFQSTVSKF---DAMPEALKSLIFPNFELPHRFTNF-L 56

DB 210 iatealliterayvyrlnldqvfycklleeanrgspaeamvknkfmissinafhskfll 269

QY 57 KETEORLALWEGRSNAQIRYQIRIGDVMKLNIOGMKHLAAHLWKHSEALENGIKSSR 116

DB 270 pelekrmqewe-----ttrigdlqlaplflkmygeyvkfgdnaveivkmntervp 321

QY 117 RLENFCRDFELQKVC-YLPNTLTLRLPLRLMHYKOVLERLCKHHPSPHADPRCRAALA 175

DB 322 qfksvteeidqkicgsltqhnmlepiqripyemilkyllklsdpdpdndakksle 381

QY 176 EITEMVAQLHGTMIKMN 193

DB 382 iistaashsnsairkmen 399

XX  
AC AAM39338 standard; Protein; 523 AA.

XX  
AC AAM39338;

XX  
DT 22-OCT-2001 (first entry)

XX  
DE Human polypeptide SEQ ID NO 2483.

XX  
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;

XX  
KW peripheral nervous system; neuropathy; central nervous system; CNS;

XX  
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;

XX  
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;

XX  
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;

XX  
KW leukaemia.

XX  
OS Homo sapiens.

XX  
PN WO200153312-A1.

XX  
PD 26-JUL-2001.

XX  
PF 26-DEC-2000; 2000WO-US34263.

XX  
PR 21-JAN-2000; 2000US-0488725.

XX  
PR 25-APR-2000; 2000US-0552317.

XX  
PR 09-JUL-2000; 2000US-0598042.

XX  
PR 19-JUL-2000; 2000US-0620312.

XX  
PR 03-AUG-2000; 2000US-0653450.

XX  
PR 14-SEP-2000; 2000US-0662191.

XX  
PR 19-OCT-2000; 2000US-0693036.

XX  
PR 29-NOV-2000; 2000US-0727344.

XX  
PA (HYSE-) HYSEQ INC.

XX  
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;

XX  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

XX  
DR WPI; 2001-442253/47.

XX  
DR N-PSDB; AAI58494.

XX  
PT Novel nucleic acids and polypeptides, useful for treating disorders

XX  
PT such as central nervous system injuries -

XX  
PS Example 4; SEQ ID NO 2483; 10078pp; English.

XX  
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and

XX  
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,

XX  
CC immunosuppressant and cytostatic activity. The polynucleotides are useful

XX  
CC in gene therapy. A composition containing a polypeptide or polynucleotide

XX  
CC of the invention may be used to treat diseases of the peripheral nervous

XX  
CC system, such as peripheral nervous injuries, peripheral neuropathy and

XX  
CC localised neuropathies and central nervous system diseases, such as

XX  
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

XX  
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the

XX  
CC utilisation of the activities such as: Immune system suppression,

XX  
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

XX  
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,

XX  
CC assays for receptor activity, arthritis and inflammation, leukaemias and

XX  
CC C.N.S disorders.

XX  
CC Note: The sequence data for this patent did not form part of the printed

XX  
CC specification.

XX  
SQ Sequence 523 AA;

Query Match 17.2%; Score 176; DB 22; Length 523;

Best Local Similarity 25.0%; Pred. No. 5.6e-10;

Matches 49; Conservative 43; Mismatches 92; Indels 12; Gaps 4;

QY 1 IAKEVSTTERTYKDL-EVITSWFQSTVSKEDAMPALKSLIFPNFELPHKFTNFKEL 59

DB 114 vineimsterhyikhkldicegylkqcrkrdmfsdeqlkvifgniediyrfmgfvdrl 173

QY 60 EORLALWEGRSNAQIRYQIRIGDVMKLNIOGMKHLAAHLWKHSEALENGIKSSRLE 119

DB 174 ekdy-----nnddphlseigpcfhqdgqfwiyseycnnhldacmelsklmksdyqh 226

```

Qy 120 NF--CRDFELQKVCYPLNTFTFLRLHRLMHYKQVLERLCKHPPSHADFRCRAALAEI 177
Db 227 ffeacr--llqgmldatdgflitpvgkckypqlqaellyktaqhsdyryvaalaavm 284

Qy 178 TEMVAQLHGHTMKMEN 193
Db 285 rnvgtqgnerkrllen 300

```

RESULT 6  
AAM41124  
ID AAM41124 standard; Protein; 647 AA.

DT 22-OCT-2001 (first entry)

XX Human polypeptide SEQ ID NO 6055.

Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
peripheral nervous system; neuropathy; central nervous system; CNS;  
Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
leukemia.

OS : Homo sapiens.

XX PN WO200153312-A1.

XX  
PD  
26-JUL-2001.XX  
PF 26-DEC-2000;

21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.  
PR 09-JUL-2000; 2000US-0598042

PR 19-JUL-2000; 2000US-0620312.  
PR 03-AUG-2000; 2000US-0653450

PR 14-SEP-2000; 2000US-0662191.  
PR 19-OCT-2000; 2000US-0693036

PR 29-NOV-2000; 2000US-0727344.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;

WPI: 2001-442253/47.

DR N-PSDB; AA160280.

PT Novel nucleic acids and polypeptides, useful for treating disorders  
PT such as central nervous system injuries -

PS Example 2; SEQ ID NO 6055; 10078pp; English;

The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAM38642-AAAM42213) with neurotrophic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: Immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis and inflammation, leukaemias and C.N.S disorders.

CC Note: The sequence data for this patent did not form part of the printed CC specification.

XX  
SQ Sequence 647 AA;

Query Match	17.2%;	Score 176;	DB 22;	Length 647;
Best Local Similarity	25.0%;	Pred. No. 7.5e-10;		
Matches 49: Conservative	43;	Mismatches 92;	Indels 1	

Qy	1	IAKEVSTTERTYTKDL- EVITSWFOSTVSKEDAMPEALASLI FPNEPEPLKHFHTNLKEI	59
Db	238	vineimstehykhkldicegylkqcrkrd fmsdeqlkvifgniedlyrfqmgfvrdl	297
Qy	60	EORLALWEGRSNAQIRDYQRIGDVMLK NIOGMKHLAAHLWKHSEALEENGTKSSRRLE	119
Db	298	ekqy-----nnddphlseipcfiehqdg fwiyseycnnhldacmelslknkdsryqh	350
Qy	120	NF--CRDFELQKCYCLPLNTFLRLPL HRLMHYQVLERLCKKHPPSHADPRCRAALAEI	177
Db	351	ffeatr--llqgmidaidgfllltpqkic kypqlaeillkkytaqghsdryyvaalavm	408
Qy	178	TEMVAQLHGHTMIKMN	193
Db	409	rnvtgaqnerkrillen	424

## RESULT

AAB97025

AAB97025  
 ID AAB97025 standard; protein; 619 AA.

AA AAB97025:

XX  
DT 16-JUL-2001 (first entry)

XX Human colon carcinoma suppressor gene-related protein.

XX Human; colon carcinoma suppressor gene; cytostatic;  
KW  
KW armadillo repeat; colon cancer.

XY  
OS  
Homo sapiens.

XX  
PN JP2001057888-A.XX  
PD  
06-MAR-2001

XX 20-AUG-1999: 99.TP-0234809.

XX  
PR 20-AUG-1999: 99JP-0234809.

XX PA (DAUC) DATIHT PHARM CO LTD.

XX  
DR  
WPT: 2001-310665/33

DR WPI; 2001-310663/  
DR N-PSDB: AAF99965.

XX  
PT  
Colon carcinoma suppressor gene-related protein -

xx PS Claim 1: Fig 1: 20pp: Japanese.

ax The present sequence is a novel colon carcinoma suppressor gene-related  
cc protein. The invention relates to the present sequence having a 619  
cc residue amino acid sequence, to a sequence containing the present  
cc sequence, and to a sequence being at least 70% homologous to the  
cc present sequence. The protein binds to the armadillo repeat site  
cc of the protein encoded by the cancer inhibiting gene of colon cancer.  
cc It may be used in the treatment of colon tumours.

AA	Sequence	619 AA;
SO		

Query Match 16.7%; Score 171; DB 22; Length 619;  
Best Local Similarity 26.0%; Pred. No. 2.4e-09;

1 1AKEVSTTERTYLKDL-EVITSWFOSTVSKEDAMPEALKSLIFPNFEPLHKFHTNFLKEI 59

Db 217 vinsllsterdykhlrdicegyvrgkradmfsseqqlrtifgniediyrcqkafvkal 276  
QY 60 EORLALWEGSRNAQIRDYQIRIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLE 119  
Db 277 eqrf-----nrphlselfgacflhqdgadqglyseycnnhpnacvelrltklskyv 329  
QY 120 NF--CRDFELQKVCYLPLNTFLRLHRLMHYKQVLERLCKHHPPSHADFRDCRAALAEI 177  
Db 330 ffeacr--llqkmidisldgfltltpvqkickyplqlaelkkythpqrhdfkveaalham 387  
QY 178 TEMVAQLHGTMIKMEN 193  
Db 388 knvaqlinerkrklen 403  
RESULT 8  
AAV71160  
ID AAY71160 standard; Protein; 1683 AA.  
XX AC AAY71160;  
DT 08-SEP-2000 (first entry)  
XX DE Rat phosphodiesterase interacting protein, M14.  
XX KW Rat; phosphodiesterase interacting protein; M14; PDE; CAMP-PDE;  
KW cyclic adenosine monophosphate phosphodiesterase; antiasthmatic; asthma;  
KW antiinflammatory; antipsoriatic; dermatological; antibacterial; shock;  
KW analgesic; immunosuppressive; antitumor; vasotropic; antiarthritic;  
KW antidiabetic; urticaria; antiallergic; antiarteriosclerotic; diagnosis;  
KW antirheumatic; treatment; inflammatory disease; psoriasis; arthritis;  
KW atherosclerosis; Crohn's disease; cystic fibrosis; chronic bronchitis;  
KW eosinophilic granuloma; proliferative skin disease; ulcerative colitis;  
KW reperfusion injury; atopic dermatitis; diabetes insipidus;  
KW conjunctivitis; adult respiratory distress syndrome; allergic rhinitis;  
KW arterial restenosis; ankylosing spondylitis; transplant rejection;  
KW graft versus host disease.  
XX OS Rattus sp.  
XX PN WC200027861-A1.  
XX PD 18-MAY-2000.  
XX PF 12-NOV-1999; 99WO-US26860.  
XX PR 12-NOV-1998; 98US-0108255.  
XX PA (STRD ) UNIV LELAND STANFORD JUNIOR.  
XX PI Conti M, Pahlke G;  
XX PS WPI; 2000-376479/32.  
XX PT Polynucleotide encoding a phosphodiesterase (PDE) interacting  
PT polypeptide, useful for diagnosis and treatment of asthma, cystic  
PT fibrosis, Crohn's disease, and rheumatoid arthritis -  
XX PS Disclosure; Fig 6; 77pp; English.  
XX CC The present sequence is a phosphodiesterase (PDE) interacting protein,  
CC M14 from rat. The protein modulates the functions and properties of PDEs,  
CC specifically CAMP-PDEs, and also targets them to specific subcellular  
CC compartments. The present sequence  
CC can be used in the diagnosis and treatment of disease conditions  
CC associated with PDE activity. The diseases include asthma, cystic  
CC fibrosis, inflammatory airway disease, chronic bronchitis, eosinophilic  
CC granuloma, psoriasis, proliferative skin diseases, endotoxic shock,  
CC septic shock, ulcerative colitis, Crohn's disease, reperfusion injury,  
CC inflammatory arthritis, atopic dermatitis, urticaria, adult respiratory  
CC distress syndrome, diabetes insipidus, allergic rhinitis, allergic  
CC conjunctivitis, vernal conjunctivitis, arterial restenosis,

CC atherosclerosis, inflammatory diseases associated with irritation and  
CC pain, rheumatoid arthritis, ankylosing spondylitis, transplant  
CC rejection and graft versus host disease, disease conditions associated  
CC with hypersecretion of gastric acid, and disease conditions in which  
CC cytokines are mediators.  
XX SQ Sequence 1683 AA;  
Query Match 16.5%; Score 169; DB 21; Length 1683;  
Best Local Similarity 24.2%; Pred. No. 1.5e-08;  
Matches 48; Conservative 50; Mismatches 80; Indels 20; Gaps 5;  
QY 4 EVSTTERTYKLEVITSWFQSTVSKEDAMPEALKSLIFPNFPLHKHFNFLKTEORL 63  
Db 1202 elieteerymddqlvievfqkrmaesgfiteaemalifvwnkelmsntkllkairv- 1260  
QY 64 ALWEGSRNAQIRDYQIRIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLENFCR 123  
Db 1261 ----kktggekmpvenmgdilaee---lshmqyirfcscql----ngaallqktdeda 1309  
QY 124 DFE--LQKVCY-----LPLNTELLRLHRLMHYKQVLERLCKHHPPSHADFRDCRAALA 175  
Db 1310 dfkeflkklasdpckgmplssflkpmqtrtryllirsilentpqnhvdhsslkiale 1369  
QY 176 EITEMVAQLHGTMIKMEN 193  
Db 1370 raeelcsqynegvreklen 1387  
RESULT 9  
AAV7450  
ID AAY57450 standard; Protein; 1658 AA.  
XX AC AAY57450;  
DT 28-FEB-2000 (first entry)  
XX DE Mouse Ese2L protein sequence.  
XX KW Mouse; murine; Ese2; endocytosis; vesicular trafficking;  
KW regulation; actin cytoskeleton; detection; cancer; infection;  
KW EH-domain and SH3-domain regulator of endocytosis; anticancer;  
KW antiproliferative; antiviral.  
XX OS Mus sp.  
XX PN WO9955728-A2.  
XX PD 04-NOV-1999.  
XX PF 27-APR-1999; 99WO-CA00375.  
XX PR 27-APR-1998; 98CA-2230201.  
XX PR 05-FEB-1999; 99US-0118739.  
XX PA (HSCR-) HSC RES & DEV LP.  
XX PI Egan SE, Wang W, Sengar A;  
XX PS WPI; 2000-052802/04.  
XX DR N-PSDB; AAZ39026, AAZ39027.  
XX PT New nucleic acid encoding EseL and 2 proteins, involved in regulation  
PT of endocytosis, used e.g. for treating cancer or preventing viral  
PT infection -  
XX PS Claim 33; Page 69-70; 99pp; English.  
XX CC The present invention specifically describes mammalian EseL and 2  
CC proteins (I) and their splice variants (Ese = EH-domain and SH3-domain  
CC regulator of endocytosis). (I) are involved in regulation of clathrin-  
CC mediated endocytosis (as a complex with Esp15 protein), vesicular









```
CC produce recombinant TRIO polypeptides, e.g. for antibody production
CC or screening for modulators of activity. Labelled antibodies that
CC bind TRIO specifically can be used to detect/quantify TRIO activity,
CC particularly for diagnosis and phenotyping of neoplastic or
CC hyperplastic disease. Modulators of TRIO activity (e.g. antisense
CC nucleic acids, antibodies, peptides or mimetics) can be used to
CC reorganise the actin cytoskeleton (claimed), e.g. in cases of wound
CC healing and/or tumour metastasis, to treat an oncogene (claimed),
CC or more generally to control growth, differentiation, migration
CC and/or survival of cells, e.g. regulation of the immune response to
CC infection, treatment of impaired immune response (as in chronic
CC granulomatous disease), control of apoptosis in cancer therapy, and
CC treatment of degenerative diseases (e.g. Parkinson's, Alzheimer's
CC or Huntington's, amyotrophic lateral sclerosis, gastric ulcers,
CC Wilm's tumour etc.). Transgenic animals can be used as models to
CC characterise TRIO genes and proteins.
XX
SQ Sequence 2861 AA;
Query Match 13.5%; Score 138; DB 18; Length 2861;
Best Local Similarity 22.5%; Pred. No. 5.5e-05;
Matches 41; Conservative 47; Mismatches 64; Indels 30; Gaps 7;
QY 1 IAEVSTTERTYTKDLEVTISWFQSTVSKEDAMPEALK---SLIFNFEPLHKEFTN-FL 56
Db 1914 vlqelveterdyvrdlgyvvgygma-lmkedgvpdmkgdkivfgnhiqygdwhrdfl 1972
QY 57 KELEORLALWEGRSNAQIRDYQRIGDVMKNIQGMKHLAAHLWKHSEALENGIKS-- 114
Db 1973 gelekl-----edpekigslfvkherrl-----hmy-----laycnpkpkseh 2011
QY 115 --SRLENFCRDFELQKVCVYLPLNTFLRLPLHRLMHYKQVLERLCKRHPPSHADRCRA 172
Db 2012 ivseyidtfedkqlqgrhrlqldtlkpvqrimkyqlllkflkyskksldtselel 2071
QY 173 AL 174
Db 2072 av 2073
RESULT 15
AAW64468
ID RAW64468 standard; Protein; 1461 AA.
AC AAW64468;
XX
XX 16-OCT-1998 (first entry)
XX Human secreted protein from clone CW420_2.
DE
DE Secreted protein; nutrition; cytokine; cell proliferation; activin;
KW differentiation; immune system; stimulator; suppressor; tissue growth;
KW haematopoiesis regulation; inhibin; chemotactic; chemokinetic; cadherin;
KW haemostatic; thrombolysis; receptor; ligand; anti-inflammatory.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 364 /label= unknown
FT Misc-difference 369 /label= unknown
FT Misc-difference 1433 /label= unknown
XX
XX W09830589-A2.
PN
XX
XX 16-JUL-1998.
XX
XX 20-DEC-1997; 97WO-US23506.
XX
XX 18-DEC-1997; 97US-0993228.
```

```
PR 10-JAN-1997; 97US-0781225.
XX (GEMY ) GENETICS INST INC.
XX Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
XX WPI; 1998-413682/35.
DR N-PSDB; AAV46316.
XX
XX New isolated nucleic acids and secreted proteins - isolated from
PT human foetal kidney, adult placenta, adult colon, adult testes,
PT foetal brain and adult brain cDNA libraries
XX
PS Claim 24a; Page 75-80; 122pp; English.
XX
XX This sequence represents a novel secreted protein from clone CW420_2
CC isolated from a human fetal brain cDNA library. This novel protein may
CC have biological activity, e.g. nutritional activity, cytokine and cell
CC proliferation/differentiation activity, immune stimulating or suppressing
CC activity, haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity or other activities.
XX
SQ Sequence 1461 AA;
Query Match 12.8%; Score 131; DB 19; Length 1461;
Best Local Similarity 22.7%; Pred. No. 0.00012;
Matches 47; Conservative 43; Mismatches 97; Indels 20; Gaps 8;
QY 1 IAEVSTTERTYTKDLEVTISWFQSTVSKEDAMPEALKSLIFNFEPLHKEFTN-FLKEIE 60
Db 623 vinelfterahrvtlkvidqvfyqrvsregilspselrkifsnledilqlhig----in 678
QY 61 ORLALWEGRSNAQIRDYQRIGDVMKNIQGMKHLAAHLWKHSE-ALEALENGIKS 114
Db 679 eqmkavrkknetsvid--qigedlltwfsgpgeeklkhaaatfcnqpfalemiksrqkk 736
QY 115 SRLENFCRDFELQKVCVYLPLNTFLRLPLHRLMHYKQVLERLCKH--HPPSHADFR--- 168
Db 737 dsrfgtfgvqdaesnpicrrllqldkdiptgmqrtrkypllldniakytewpterekvkaa 796
QY 169 -DCRAALAEITEMVAQLHGTMRKWFN 194
Db 797 dhcrqilnyvngavkkaenkq-rledy 822
Search completed: December 6, 2001, 08:51:43
Job time: 335 sec
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:55:09 ; Search time 36.72 Seconds  
(without alignments)  
118.890 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_544\_737  
Perfect score: 1023  
Sequence: 1 IAKEVSTERTYTKDLEVT.....AEITEMVAQLHGTMIKMF 194

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	138	13.5	761	4	US-09-625-188-14
2	138	13.5	2860	2	US-08-826-267-2
3	131	12.8	1461	2	US-08-993-228-10
4	120	11.7	1244	3	US-09-356-952-7
5	118	11.5	477	1	US-08-191-338A-2
6	94.5	9.2	912	4	US-08-943-768-2
7	86.5	8.5	626	4	US-09-155-770-7
8	80	7.8	844	1	US-07-646-537B-2
9	78.5	7.7	1724	2	US-08-477-451-15
10	77.5	7.6	583	2	US-08-616-392C-4
11	77.5	7.6	645	1	US-08-785-430-2
12	77.5	7.6	645	2	US-08-996-800-2
13	76.5	7.5	2154	2	US-08-841-349-4
14	76	7.4	534	2	US-08-878-563A-1
15	76	7.4	534	4	US-09-270-117-1
16	75.5	7.4	748	3	US-08-904-871-6
17	75.5	7.4	748	3	US-08-904-871-13
18	75	7.3	521	2	US-08-878-563A-3
19	75	7.3	521	4	US-09-270-117-3
20	74	7.2	584	2	US-08-415-593-41
21	74	7.2	1143	2	US-08-310-912A-108
22	74	7.2	1143	5	PCT-US95-04589-108
23	74	7.2	1143	5	PCT-US95-04589-108
24	74	7.2	1144	1	US-08-261-663A-2
25	74	7.2	1144	1	US-08-261-663A-4
26	74	7.2	1144	3	US-08-930-996A-9
27	74	7.2	1144	5	PCT-US95-07754A-2

28	74	7.2	1144	5	PCT-US95-07754A-4
29	73.5	7.2	487	2	US-08-724-394A-7
30	72.5	7.1	871	2	US-08-775-009-34
31	72.5	7.1	871	2	US-08-775-009-35
32	71.5	7.0	732	1	US-08-481-626-2
33	71.5	7.0	732	4	US-08-989-299-4
34	71.5	7.0	1306	4	US-08-989-299-7
35	71	6.9	326	2	US-08-306-511A-10
36	71	6.9	326	2	US-08-893-274-10
37	71	6.9	326	3	US-08-581-918A-10
38	71	6.9	326	4	US-08-346-147B-10
39	71	6.9	326	4	US-08-822-936-10
40	71	6.9	326	5	PCT-US95-04636-10
41	71	6.9	1911	1	US-08-348-006B-5
42	71	6.9	1911	2	US-08-800-825A-5
43	71	6.9	1911	4	US-09-158-657-5
44	71	6.9	1911	5	PCT-US94-10166-5
45	70.5	6.9	297	2	US-09-027-013-3

ALIGNMENTS

RESULT 1  
US-09-625-188-14  
; Sequence 14, Application US/09625188  
; Patent No. 6307037  
; GENERAL INFORMATION:  
; APPLICANT: No. 6307037artis AG  
; TITLE OF INVENTION: Fungal Target Genes and Methods  
; FILE REFERENCE: PB/5-31285p1  
; CURRENT APPLICATION NUMBER: US/09/625,188  
; CURRENT FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 44  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 14  
; LENGTH: 761  
; TYPE: PRT  
; ORGANISM: Ashbya gossypii  
US-09-625-188-14

Query Match	13.5%	Score 138;	DB 4;	Length 761;
Best Local Similarity	26.2%	Pred. No. 1.5e-07;		
Matches	56;	Conservative 37;	Mismatches 83;	Indels 38; Gaps 9;
QY	1	IAKEVSTERTYTKDLEVTSTVSKEDAMPEALKSLIFNFEPLHFKHFTNFKLEI	60	
Db	267	IVKEFVETERYVHDLEVISKYRQQLLENIISSEEL-YMLFFNLNEIDFQRRFLVALE	325	
QY	61	QRLALMEGRSNAQI-RDYQRIGDVMLKNIQGMKHLAAHLKMKHSEALENGIKSS-RRL	118	
Db	326	-----INGQVPAQAQRIGALFMHS-----KHFFKLYEPWSIGQNAAINFISSSPDKM	372	
QY	119	ENCFRPELQKVCYLPNTFLPLRLHRLMYKQVLERLCK--HHPSPHADFRCDRAAL--	174	
Db	373	QS--QEFVIGN--KMELQSFLLKPQVRLCRYPLLLKDLKLSVKTKSDVDTKQLTALEI	428	
QY	175	-----AEITEMVAQLHGTMIKMF	194	
Db	429	SKSIARSINENQRRTENHEVKKLYGRVNNKGY	462	
RESULT 2				
US-08-826-267-2				
; Sequence 2, Application US/08826267				
; Patent No. 5994070				
; GENERAL INFORMATION:				
; APPLICANT: Streuli, Michel				
; TITLE OF INVENTION: No. 5994070el TRIO Molecules and Uses Related Thereto				
; NUMBER OF SEQUENCES: 2				
; CORRESPONDENCE ADDRESS:				
; ADDRESSEE: LAHIVE & COCKFIELD				

Query Match 12.8%; Score 131; DB 2; Length 1461;  
Best Local Similarity 22.7%; Pred. No. 2.7e-06;

Qy	60	1	IAKEVSTERTYIKDLLEVITSWFQSGVTSKEDAMPEALKSLIPNFNPLPHKFTNFLEKETE
Db	623	623	VINELFYTERAHVRTLKVLDQVFYQVRSREGILSPSELKFSNLEDILQLHTG----LN 678
Qy	61	61	QRLALWEGRSNAQIRDYQIRGDVYMLNIQ-----MKHLAAHLWKHISE-ALEALENGIKS 114
Db	679	679	EQMKAVKRNKNETSVID--QIGEDLLTWFGSGPEKLUKHAATAFCSNOPALEMIKSRQKK 736
Qy	115	115	SRLENFCRDFELQKVC-YLPLTFTELLRPLRHLMYKQVLERLCKH--HPPSHADSR--- 168
Db	737	737	DSRFQFVQDAESNPLCRLRLKDIIPTQOMRLTKYPLLLDNLIAKYTENPTREKVKKAA 796
Qy	169	169	-DCRAALAEITEMVAQLHGTMIKMENF 194
Db	797	797	DHCROILNYVNOAKVKAENKQ--RLEDY 822

```

RESULT      4
US-09-356-952-7
; Sequence 7, Application US/09356952
; Patent No. 6117663
; GENERAL INFORMATION:
; APPLICANT: Boriack-Sjodin, Ann
; APPLICANT: Margarit, S. M.
; APPLICANT: Bor-Sogi, Dafna
; APPLICANT: Cole, Philip
; APPLICANT: Kuriyan, John
; TITLE OF INVENTION: A CRYSTAL OF A RAS-SOS COMPLEX AND METHODS OF USE
; FILE OF INVENTION: THEREOF
; FILE REFERENCE: 600-1-228N
; CURRENT APPLICATION NUMBER: US/09/356,952
; CURRENT FILING DATE: 1999-07-19
; EARLIER APPLICATION NUMBER: 60/093,631
; EARLIER FILING DATE: 1998-07-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1244
; TYPE: PRT
; ORGANISM: Rattus rattus

```

US-09-356-952-7

Query Match 11.7%; Score 120; DB 3; Length 1244;  
Best Local Similarity 22.1%; Pred. No. 4.2e-05;  
Matches 44; Conservative 33; Mismatches 90; Indels 32; Gaps 5;  
QY 9 ERTYKDLVITSWF-----QSTVSKEDAMPEALKSLIFPNFEPLHKHFTNFKETEQR 63  
DB 252 EAEYVQQLHILVNNFLRPLRMAASSKKPPIITHDDVSSIFLNSSETIMFLHQIFVQGLKARI 311  
QY 64 ALWEGRSNAIRD-----XORIGDVMKNIQGMKHLAAHLWKHSEALEALENGIKSS 115  
DB 312 ASWPTLVLADELFDILLPMUNIQE-----FVRNHQVSLQILAH-----CKON 353  
QY 116 RLLENFCRDFELQKVC-YLPLNTFLRLPLRLMHLMYKQVLERLCKHPPSPHADPRDRAAL 174  
DB 354 RQFDKLLKQYEAKPDCEERTLETFELTYPFQIPRYITLTHELLAHTPHEHVERNSLDYAK 413  
QY 175 AEITEMVAOLHGTMIKEN 193  
DB 414 SKLELSRVMHDEVSTEN 432

RESULT 5

US-08-191-338A-2  
; Sequence 2, Application US/08191338A  
; Patent No. 5763164  
; GENERAL INFORMATION:  
; APPLICANT: Calenoff, Emanuel  
; TITLE OF INVENTION: Immunogenic Cancer Proteins and Peptides  
; TITLE OF INVENTION: and Method of Use  
; NUMBER OF SEQUENCES: 38  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Arnold, White & Durkee  
; STREET: 321 No. 5763164th Clark Street, Suite 800  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: USA  
; ZIP: 60610  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/191,338A  
; FILING DATE: 12-JAN-1994  
; CLASSIFICATION: 436  
; ATTORNEY/AGENT INFORMATION:  
; NAME: No. 5763164thrup, Thomas E.  
; REGISTRATION NUMBER: 33,268  
; REFERENCE/DOCKET NUMBER: nwn:002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 312-744-0090  
; TELEFAX: 312-755-4489  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 477 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-191-338A-2

Query Match 11.5%; Score 118; DB 1; Length 477;  
Best Local Similarity 21.0%; Pred. No. 1.7e-05;  
Matches 44; Conservative 45; Mismatches 67; Indels 54; Gaps 9;  
QY 1 IAKEVSTTERTYKDLVITSWFQSTVSKEDAMPEAL---DAMPEAL---KSLIFPNFEPLHKHFTN 54  
DB 52 VLNELIQTERVYVRELYTVLLGYRAEMDNPEMFDLPPLLRNKKDILFCGMAEIVEFHND 111

QY 55 FLKEIDQRLALWEGRSNAQIRDYQVIGDVMKNIQGMKHLAAH-----LW-KHSEA 104  
DB 112 IF-----LSSLENCAP-----ERVGPCFLERKDDDFQYAKYCONKPRSETIWRKYSEC 161  
QY 105 LEALENGIKSSRRLENFCRDFELQKVCYLPNTFLRLPLRLMHLMYKQVLERLCKHPPSPH 164  
DB 162 AFQECQCRKLKHLR-----LDSYLLKPVQIRITKYQLLLKELLKYS----- 202  
QY 165 ADPRDC-----RAALAEITEMVAOLHGTM 188  
DB 203 ---KCEGSALLKALDAMDLLKSYNSDM 229

RESULT 6

US-08-943-768-2  
; Sequence 2, Application US/08943768C  
; Patent No. 6238881  
; GENERAL INFORMATION:  
; APPLICANT: Hart, Matthew J.  
; TITLE OF INVENTION: No. 6238881el Nucleic Acids and Polypeptides Related to a  
; TITLE OF INVENTION: Guanine Exchange Factor of RHO GTPase  
; FILE REFERENCE: 1023-US  
; CURRENT APPLICATION NUMBER: US/08/943,768C  
; CURRENT FILING DATE: 1997-10-06  
; EARLIER APPLICATION NUMBER: 60/029,979  
; EARLIER FILING DATE: 1996-11-06  
; NUMBER OF SEQ ID NOS: 12  
; SOFTWARE: Patent in Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 912  
; TYPE: PRT  
; ORGANISM: Human p115 GEF-Rho  
US-08-943-768-2

Query Match 9.2%; Score 94.5; DB 4; Length 912;  
Best Local Similarity 22.0%; Pred. No. 0.027;  
Matches 47; Conservative 45; Mismatches 87; Indels 35; Gaps 10;  
QY 1 IAKEVSTTERTYKDLVITSWFQSTVSKEDAMP-EALKSLIFPNFEPLHKHFTNFKEL 59  
DB 420 VISELLVTEAAHVMDLVLDLFFQPMACCLFPPLBELQN-IPPSLDELIEVLSFLDRL 478  
QY 60 EQRLALWEGRSNAQIRDY---QRIGDVMKNIQGMKHLAAHLW-----KHSEALEA 107  
DB 479 MKR-----RQESGYLIEEIGDVLARFDG---AEGSWFKISSRFSRQSFALQEQ 525  
QY 108 LENGIKSSRRLENFCRDFELQKVC-YLPLNTFLRLPLRLMHLMYKQVLERLCKH-HPPSHA 165  
DB 526 LKAKQRKDPFCFAVQEAESRPRCRRLQLKDMIPTQMRLTKYPLLLQSIGQNTSEPTER 585  
QY 166 DFRD-----CRAALAEITEMVAOLHGTMIKENF 194  
DB 586 EKVELAAECCREILHHVNOAVRDME-DLLRLKDY 618

RESULT 7

US-09-155-770-7  
; Sequence 7, Application US/09155770A  
; Patent No. 6300484  
; GENERAL INFORMATION:  
; APPLICANT: Duhl, David  
; TITLE OF INVENTION: DNA ENCODING DP-75 AND A PROCESS FOR ITS USE  
; FILE REFERENCE: 200130.418  
; CURRENT APPLICATION NUMBER: US/09/155,770A  
; CURRENT FILING DATE: 1998-09-30  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7  
; LENGTH: 626  
; TYPE: PRT  
; ORGANISM: Homo sapien

Qy 63 LA 6

; TITLE OF INVENTION

:	TITLE OF INVENTION:	NOVEL POLYNUCLEOTIDES PANCA1A
:	TITLE OF INVENTION:	AND PANCLB ASSOCIATED WITH PANCREATIC CANCER

Query Match 7.6%; Score 77.5; DB 1; Length 645;  
Best Local Similarity 20.8%; Pred. No. 1.6;  
Matches 51: Conservative 32; Mismatches 85; Indels 77; Gaps 10;

Qy	62	RLALWGRSNAQIRDYQRIGDVMLKNIQG-----MKHLAAHLWKHSEALEA-----	107
Db	201	AGAYWRGDSN-----NKLQRIYGTAFDPKKELKAHLQMLERKEROHRKTIKGL	250
Qy	108	-----LENGIKSSRLENFCRDFELQ---KVCYCLPL--NTFELLRLPLHRL	146
Db	251	ELFTNSOLVGAGLPLWLPNGATTIRREIERYIVDKVSMGYDVHYITPVLANVDLYKTSGHW	310
Qy	147	MHYKQ-----VLERL-CKHHPPSHAD---FRDCRAALAEITEMVAOLHGT	187
Db	311	DHQEDMFPPQMLDETSWVLRPWCNCPHHMIIYANKPHSVRELPIRIAE-----GT	362
Qy	188	MIRME	192
Db	363	MHYRE	367

```

RESULT 12
US-08-996-800-2
; Sequence 2, Application US/08996800
; Patent No. 5928920
; GENERAL INFORMATION:
; APPLICANT: Hodgson, John
; APPLICANT: Lawlor, Elizabeth
; TITLE OF INVENTION: No. 5928920el tRNA synthetase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
;

```

```

; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/996,800
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/785,430
; FILING DATE: 17-JAN-1997
; APPLICATION NUMBER: 9601096.2
; FILING DATE: 19-JAN-1996
; APPLICATION NUMBER: 9622617.0
; FILING DATE: 30-OCT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Gimmi, Edward R.
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P31354-6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 645 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-996-800-2

Query Match 7.6%; Score 77.5; DB 2; Length 645;
Best Local Similarity 20.8%; Pred. No. 1.6;
Matches 51; Conservative 32; Mismatches 85; Indels 77; Gaps 10;

QY 2 AKEVSTERTYRKDLEIVTSWFSQTSKEDAMPEALKSLIFPNFPLKHFHTNFKIEQ 61
Db 146 AKELFNDXEYKLELIDAIPEVDENVTLYSQDFTDLGRGVHVPSTAKIKEP-----KLLST 200
QY 62 RLALWEGRSNAQRDQYRIGDVMLKNTQG-----MKHLAHLWKHSEALEA----- 107
Db 201 AGAYWRGDSN-----NKMLRIYGFAPFDKKELKAHLQMLEERKDRHRKIGKEL 250
QY 108 -----LENGIKSSRRLENFCRDFELQ---KVCYLPL--NTFLLRPLHRL 146
Db 251 ELFTNSOLVCAGLPLWLPNGATIRREIRYIVDKESMGVDHYVTPVLNVLDLYKTSQHW 310
QY 147 MHYKQ-----VLERL-CKHHPPSHAD-----FRDCRAALAEITEMVAQLHGT 187
Db 311 DHYQEDMFPQMQLDETESMVLRPWNCPPHMMIYANKPHSYRELPIRIAEI-----GT 362
QY 188 MIKWE 192
Db 363 MHRYE 367

RESULT 13
US-08-841-349-4
; Sequence 4, Application US/08841349B
; Patent No. 5955594
; GENERAL INFORMATION:
; APPLICANT: MISHRA, LOPA
; TITLE OF INVENTION: GENES CODING PROTEINS FOR EARLY LIVER DEVELOPMENT.
; FILE REFERENCE: XX/PO4470US0
; CURRENT APPLICATION NUMBER: US/08/841,349B
; CURRENT FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 2154
; TYPE: PRT
; ORGANISM: Mus musculus

```

```

US-08-841-349-4

Query Match 7.5%; Score 76.5; DB 2; Length 2154;
Best Local Similarity 24.1%; Pred. No. 13;
Matches 51; Conservative 29; Mismatches 61; Indels 71; Gaps 14;

QY 3 KEVSTTE---RTYKLDLEIVTSWF---QSTVSKEDAMPEALKSLIFPNFPLKHFHTNF 55
Db 1041 REASLGASKLQQLRDLDDFQSWLSRTQTAIASED--MPNTLTEA-----EKLITQHENI 1094
QY 56 LKEIQRLALWEGRSNAQRDQYRIGDVMLKNTQGKMKHLAHLWKHSEALEALENGIKSS 115
Db 1095 KNEIDN---YE-----EDYQKMRDMGEMVTQGGTD-AQYMFRLR-QRLQALDTG---- 1137
QY 116 RLLENFCROFELQKVCYLPLNTLRLPLHRLMHYKQVLERLCKHHPPSHADF-RDCRAAL 174
Db 1138 -----WNLHKMWENRNL--LSQSH--AYOOFLRDTKQAE 1169
QY 175 AEI-----TEMVAQLHG---TMIKMENF 194
Db 1170 AFLNQYVLAHTEMPTTLEGAAAIKKQEDF 1201

RESULT 14
US-08-878-563A-1
; Sequence 1, Application US/08878563A
; Patent No. 5891674
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,563A
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0323 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 534 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRSTNOT04
; CLONE: 918158
; US-08-878-563A-1

Query Match 7.4%; Score 76; DB 2; Length 534;
Best Local Similarity 24.8%; Pred. No. 1.8;

```



Job time: 293 sec

Qy	46	EPLHKHTNFKLEIQEPALWEGSRNAQIDYQ---	RTGDV-----MLKNIQM	91
		: : : : : :	: : : : : :	
Db	92	EMLKFHNELTQEQVEDLSRYLSAALKUYTEQRSGDALDQCQELKLRKSOGS	151	
		: : : : : :	: : : : : :	
Qy	92	KHLAAHLWKHSEALELENIGKSRLENFCRD-----FELQKVCYL	133	
		: : : : : :	: : : : : :	
Db	152	KNPQKYSDKLOQYIDAISN---KQGELENVVDGYKTALTEECRRCFEL	197	
		: : : : : :	: : : : : :	

```

RESULT 15
US-09-270-117-1
; Sequence 1, Application US/09270117
; Patent No. 6285550
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; TITLE OF INVENTION: INSULIN RECEPTOR TYROSINE KINASE SUBSTRATE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/270,117
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/878,563

```

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36.749  
REFERENCE/DOCKET NUMBER: PF-0323 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-855-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 534 amino acids  
TYPE: amino acid  
STRAINEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: BRSTNOT04  
CLONE: 918158  
US-09-270-117-1

```

Query Match      7.4%; Score 76; DB 4; Length 534;
Best Local Similarity 24.8%; Pred. No. 1.8;
Matches 27; Conservative 18; Mismatches 40; Indels 24; Gaps

QY 46 EPLHKHTNPLEKEIORLALWEGRSNAQTRDQ-----RTGDV-----MLKNIQGM 91
      | | | | | : | | | : | | | | | : | | |
Db 92 EMLKSPHNELLTQEQVELDSRYLSAALKKYQTEQRSKDGLDKQAECLKLRKKSQS 151
      | | | | | : | | | : | | | | | : | | |

QY 92 KHLAAHLWKHSEALEALENCIKSSRLENFCRD-----FELQKVCYL 133
      | : | : | : | : | : | : | : | : | : | : |
Db 152 KNPQKYSDKELQYIDAISN---KQGLENYVSDGYKALTETECRCFFL 197
      | : | : | : | : | : | : | : | : | : | : |

```

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:52:37 ; Search time 44.17 Seconds  
(without alignments)  
334.568 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_544\_737

Perfect score: 1023

Sequence: 1 IAKEVSTTERTYKDLKLEVT.....AEITEMVAQLHGTMIKMF 194

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR:68:\*

1: pir1.\*

2: pir2.\*

3: pir3.\*

4: pir4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1023	100.0	1045	2 JC5795	CDEP protein - hum
2	190.5	18.6	875	2 T19678	hypothetical prote
3	171	16.7	496	2 T46356	hypothetical prote
4	166.5	16.3	1335	2 T18289	racGAP protein - s
5	157.5	15.4	1737	2 A92235	unconventional myo
6	154	15.1	961	2 A5380	faciogenital dyspl
7	129.5	12.7	573	2 I49342	hypothetical onco
8	123	12.0	646	2 JC5583	85K SH3 domain-con
9	122	11.9	1189	2 T42726	guanine nucleotide
10	121.5	11.9	1158	2 T50454	probable rhol GDP-
11	120	11.7	1244	2 S29083	guanine-nucleotide
12	119	11.6	552	2 T34428	hypothetical prote
13	119	11.6	2488	2 T42739	guanine nucleotide
14	118	11.5	478	4 TVHUBD	transforming prote
15	118	11.5	925	1 TVHUBD	transforming prote
16	117.5	11.5	738	2 S32372	transforming prote
17	117.5	11.5	2044	2 T13704	still life protein
18	117.5	11.5	2064	2 T13707	still life protein
19	116	11.3	872	2 T37789	Scd1 protein - fis
20	116	11.3	1260	2 S28407	guanine nucleotide
21	114.5	11.2	1334	2 T41524	rhof gdp-ntp excha
22	114	11.1	567	2 S10138	mcf2 protein - hum
23	114	11.1	736	2 A27477	cell division cont
24	113.5	11.1	620	2 T16657	hypothetical prote
25	112	10.9	685	2 JC6331	rho-type guanine e
26	111	10.9	872	2 S51620	Ost oncogene - rat
27	111	10.9	1275	2 A38985	nucleotide exchang
28	110	10.8	677	2 T15242	hypothetical prote
29	108	10.6	1275	2 T41523	hypothetical rhof

30 106.5 10.4 278 2 A60195 transforming prote  
31 105.5 10.3 1693 2 T30867 Rho-guanine nucleo  
32 105 10.3 914 2 T17233 hypothetical prote  
33 105 10.3 1271 1 TVHUBR bcr (breakpoint cl  
34 104 10.2 1591 2 A54146 invasion-inducing  
35 103.5 10.1 860 2 T15778 hypothetical prote  
36 103.5 10.1 919 2 T21663 hypothetical prote  
37 103 10.1 878 2 T51940 gene VAV2 protein  
38 101 9.9 462 2 T25544 hypothetical prote  
39 101 9.9 519 2 T38402 guanine nucleotide  
40 98 9.6 460 2 G01210 guanine nucleotide  
41 98 9.6 1899 2 T32732 PAM C-terminal int  
42 98 9.6 1919 2 T42098 PAM interacting pr  
43 97.5 9.5 862 2 T34342 hypothetical prote  
44 96.5 9.4 2137 1 SJHUB spectrin beta chai  
45 95.5 9.3 693 1 TVHUA2 bcr (breakpoint cl

## ALIGNMENTS

RESULT 1  
JC5795  
CDEP protein - human  
C:Species: Homo sapiens (man)  
C:Date: 24-Jan-1998 #sequence\_revision 13-Mar-1998 #text\_change 16-Jul-1999  
C:Accession: JC5795  
R:Koyano, Y.; Kawamoto, T.; Shen, M.; Yan, W.; Noshiro, M.; Fujii, K.; Kato, Y.  
Biochem. Biophys. Res. Commun. 241, 369-375, 1997  
A:Title: Molecular cloning and characterization of CDEP, a novel human protein contain  
ing factors.  
A:Reference number: JC5795; MUID:98086358  
A:Accession: JC5795  
A:Molecule type: mRNA  
A:Residues: 1-1045 <KOY>  
A:Cross-references: DBJ:AB008430  
C:Comment: this protein is involved in the adhesion, proliferation, and differentiati  
C:Superfamily: pleckstrin repeat homology; protein 4.1 membrane-binding domain homolo  
F:1-374/Domain: ezrin-like #status predicted <E2>  
F:42-316/Domain: protein 4.1 membrane-binding domain homology <B41>  
F:931-1027/Domain: pleckstrin repeat homology <PLK>

Query Match 100.0%; Score 1023; DB 2; Length 1045;  
Best Local Similarity 100.0%; Pred. No. 9.3e-80;  
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 IAKEVSTTERTYKDLKLEVTSNFQSTVSKEDAMPEALKSLIFPNFPLHKFTNFKLEIE 60  
Db 544 IAKEVSTTERTYKDLKLEVTSNFQSTVSKEDAMPEALKSLIFPNFPLHKFTNFKLEIE 603  
QY 61 QRLALWEGRSNAQIRDYQYRIGDVMLKNIQGMKHLAAHLKWKHSEALEENGKSSRRLN 120  
Db 604 QRLALWEGRSNAQIRDYQYRIGDVMLKNIQGMKHLAAHLKWKHSEALEENGKSSRRLN 663  
QY 121 FCRDFELQKVCYLPNTFLRLPLRLMHYKQVLERLCKHHPPSHADFRCRAALAEITEM 180  
Db 664 FCRDFELQKVCYLPNTFLRLPLRLMHYKQVLERLCKHHPPSHADFRCRAALAEITEM 723  
QY 181 VAQLHGTMIKMF 194  
Db 724 VAQLHGTMIKMF 737

RESULT 2  
T19678  
hypothetical protein C33D9.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T19678  
R:Lloyd, C.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: Z19162

```

Query Match      16.7%; Score 171; DB 2; Length 496;
Best Local Similarity 26.0%; Pred. No. 4e-07;
Matches 51; Conservative 37; Mismatches 96; Indels 12; Gaps 4;

QY 1 IAKEVSTERTYKDL-EVITSWFQSVSKEDAMPEALKSLIFPNFPLKHFHFNFLKEI 59
   : ||: |||| :|| : : : : || : || : || : || : || : || : || :
Db 94 VINEILSTERDYIKHLKDICEGYVQRCRKADMSEQLRTIFQGNIEDIVYRCQAFVKAL 153
   : || : || : || : || : || : || : || : || : || : || : || : || :
QY 60 EQRALMEGRSNAQIRDYQYIGDVMKLNIOGMKHLAAHLMKHSLEALENGIKSSRLE 119
   ||| : || : || : || : || : || : || : || : || : || : || : || :
Db 154 EQRE-----NRERPHLSLGACFLEHQADFQIYSEVCNNHPNACVLSRLTKLSKVY 206
   : || : || : || : || : || : || : || : || : || : || : || : || :
QY 120 NF--CRDFELQKVCYLPINTFLLRPLRLMHYKQVLEELCKHHPPSHADPDCRAALAEI 177
   ||| |||| : || : || : || : || : || : || : || : || : || : || :
Db 207 FFEACR--LQMKIDISLDGFLTPVOKICKYPLQLAELLYKTHPQHRDRFDVEAALHAM 264
   : || : || : || : || : || : || : || : || : || : || : || : || :
QY 178 TEMVAQLHGTMIKMEN 193
   : || : || : || : || : || : || : || : || : || : || : || : || :
Db 265 KNVAQLINERKRRLEN 280
   : || : || : || : || : || : || : || : || : || : || : || : || :

```

**RESULT**      4  
Tl8289  
C:Accession: Tl8289  
C:Species: Dictyostelium discoideum  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jul-2000  
J: Ludbrook, S.B.; Eccleston, J.; Strom, M.  
R. Biol. Chem. 272, 15682-15685, 1997  
A:Title: Cloning of a rhoGAP homolog from Dictyostelium discoideum.  
A:Reference number: Z18858; MUID: 97332648  
A:Accession: Tl8289  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1335 <LUD>  
A:Cross-references: EMBL:Y10159; NID:g2190354; PIDN:CAA71241.1; PID:g2190354  
C:Genetics:  
A:Gene: racGAP  
A:Introns: 113/1; 205/3

Query Match                  16.3%; Score 166.5; DB 2; Length 1335;  
Best Local Similarity        27.6%; Pred. No. 3.2e-06;  
Matches 59; Conservative 34; Mismatches 74; Indels 47; Gaps

Qy    2   AKEVSTERYTLKDLVIYSWFQSTVSK-----EDAMPEALKSLIFNPFPPLIKFH 52  
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db    795   AEETLTAEQTYYVKLTWVDNFTEPYKKTKQKGHLTGEEFMD-----IFNCLEVLISSH 848

Qy    53   -TNFLKEIQRALWEGRSNAQIRDYRIDGYMLKNIOGMKHLAAHIAHWKHSEALEANG 111  
   || || || || || :  
Db    849   KTNLLKPIEDRMVLWDISK-----PQMGDIFLNNTSFIR-----LYKH-----YVNYY 890

Qy    112   IKSSRRLENPCRD----FEL-----QKVCYLPINTFLRLPHRLMHYKQVLERLCKH 159  
   || || || || || :  
Db    891   DKSIRT-LNOCKEYDGFXTYNASLDYSENLSLESFLVLIPIQLPRYVMLLQDLLKY 949

Qy    160   HPSHADFDRCRAALAEITEMVAQLHGTTWKMEN 193  
   || || || :  
Db    950   TANDHEDFNOLCEALSTIKDLTESINTRKKSEDN 983

**RESULT**      5  
A59235  
unconventional myosin heavy chain MyoM - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 19-May-2000 #sequence\_revision 19-May-2000 #text\_change 08-Sep-2000

submitted to GenBank, September 1998  
A:Description: Identification of two novel and highly divergent myosins in Dictyostelium  
A:Reference number: A59235  
A:Accession: A59235  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-1737 <GEI>  
A:Cross-references: GB:AF090533; NID:g5714395; PIDN:AAD47903.1; PID:g5714396  
A:Experimental source: strain AX2  
C:Genetics:  
A:Gene: myoM  
A:Map position: 6, aldB-cabA2  
C:Superfamily: myosin motor domain homology  
F:62-874/Domain: myosin motor domain homology #status atypical <MMO>

```

Query Match      15.4%  Score 157.5;  DB 2;  Length 1737;
Best Local Similarity 25.9%  Pred. No. 2.6e-05;
Matches 49;  Conservative 37;  Mismatches 82;  Indels 21;  Gaps 7;

Qy  1  IAKEVSTTTERYLKDLVITSWFQSTVSKEDAMPALKSLFFNPPEPLHGHFTNPLKEIE 60
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db  1393  IINELIETERYDVRDLNIVVEVFLNPIREKQLLSAKDINSLSFTEIIFSINMNVLKALE 1452

Ov  61  O-RALMEGRSNAOIRDYORIGDVMYLNKIOGMKHLAAHLWKHSEALEAL-ENGIKSS--R 116

```

Db 1453 KKDPLCENIS-----VQTFELMSHYLKMWTYSCNOONALKILEEKIKNOPPR 1503  
QY 117 RLENFCRDFELQKVC-YLPNTFLRLPLRMLHYKQVLERLCKHHPPSHADRCRAALA 175  
Db 1504 EYLEFCMN---DSVCRGLPLNSFIKPKVQRICKYPLLIKETIKFTPNDDPD---KPALE 1556  
QY 176 EITEMVAQL 184  
Db 1557 EYDKKISDI 1565  
RESULT 6  
A55380  
faciogenital dysplasia-associated protein FGD1 - human  
C:Species: Homo sapiens (man)  
C:Date: 10-Feb-1995 #sequence\_revision 10-Feb-1995 #text\_change 17-Mar-1999  
C:Accession: A55380  
R:Pastoris, N.G.; Cadle, A.; Logie, L.J.; Porteous, M.E.M.; Schwartz, C.E.; Stevenson, R.  
Cell 79, 669-678, 1994  
A:Title: Isolation and characterization of the faciogenital dysplasia (Aarskog-Scott syn-  
A:Reference number: A55380; MUID:95042764  
A:Accession: A55380  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-961 <PAS>  
A:Cross-references: GB:U11690; NID:g595424; PID:g595425  
C:Superfamily: CDC24 homology; pleckstrin repeat homology  
F:373-561/Domain: CDC24 homology <CD24>  
Query Match 15.1%; Score 154; DB 2; Length 961;  
Best Local Similarity 23.7%; Pred. No. 2.5e-05;  
Matches 47; Conservative 41; Mismatches 96; Indels 14; Gaps 4;  
QY 1 IAKEVSTTERTYKLEVITTSWFQSTVSKE-----DAMPEALAKSLIFPNFELHKEHTNF- 55  
Db 377 IANELLQTEKATVSRHLDDQVFCARLLEEARNRSPADVVHGFISNICSICYFHQQPL 436  
QY 56 LKEIEORLALWEGRSNAQIRDYQRIGDVMLKNIQGMKHLAAHLWKHSEALEALENGIKSS 115  
Db 437 LPELEKREWE-----RYPRIIGDILQKLAPFLKMYGEVKNFDFRAVELVNTWTERS 488  
QY 116 RLENFCRDFELQKVC-YLPNTFLRLPLRMLHYKQVLERLCKHHPPSHADRCRAAL 174  
Db 489 TQFKVIIHEVQKEEACGNLTQHHMLEPVQIRPRYVELLKYLLKPLPHGSPDSKDAQSL 548  
QY 175 AEITEMVAQLHCTMIKME 192  
Db 549 ELIATAAEHSNAAIRKME 566

RESULT 7  
I49342  
hypothetical oncogene protein lfc [imported] - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 20-Jun-2000  
C:Accession: I49342  
R:Whitehead, I.; Kirk, H.; Togno, C.; Trigo-Gonzalez, G.; Kay, R.  
J. Biol. Chem. 270, 18388-18395, 1995  
A:Title: Expression cloning of lfc, a novel oncogene with structural similarities to gua-  
A:Reference number: A57387; MUID:95355462  
A:Accession: I49342  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-573 <RES>  
A:Cross-references: EMBL:U28495; NID:g902022; PIDN:AAC52234.1; PID:g902023  
C:Genetics:  
A:Gene: lfc  
C:Superfamily: pleckstrin repeat homology  
F:472-570/Domain: pleckstrin repeat homology <PLK>

Query Match 12.7%; Score 129.5; DB 2; Length 573;  
Best Local Similarity 21.4%; Pred. No. 0.0017;  
Matches 44; Conservative 50; Mismatches 85; Indels 27; Gaps 7;  
QY 4 EYSTTERTYKLEVITTSWFQSTVSKEADAM-DEALKSLIFFPNFELHKEHTNFLEKEI--E 60  
Db 243 ELIQTELHVRILKIMTRFLRTGMLQELQMEPEVVOGL-FPCVDELSDIHTRELNOLLER 301  
QY 61 QRLALWEGRSNAQIRDYQRIGDVMLKNIQGMKHLAAHLWKHSEALEALENGIKS 114  
Db 302 RRQALCPGSTRNFV--IHLGLDILLISQFSGSNAEQMRKTYSEFCSRHTKALKLYKELYAR 359  
QY 115 SRLENFCRDFELQKV-----CYLPLNTFLRLPLRMLHYKQVLERLCKHHPPSHAD 166  
Db 360 DKRFQFIRKMTSRSAVLKRHGVQEC-----ILLVQRTITKTPVPLINILQNSHGVEE 412  
QY 167 FRDCRAALAEITEMVAQLHGTMIKME 192  
Db 413 YODLASALGLVKELLSNVDDVHELE 438  
RESULT 8  
JC5583  
85K SH3 domain-containing proline-rich protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 23-Sep-1997 #sequence\_revision 23-Sep-1997 #text\_change 16-Dec-1998  
C:Accession: JC5583  
R:Oh, W.K.; Yoo, J.C.; Jo, D.; Song, Y.H.; Kim, M.G.; Park, D.  
Biochem. Biophys. Res. Commun. 235, 794-798, 1997  
A:Title: Cloning of a SH3 domain-containing proline-rich protein, p85SPR, and its loc-  
A:Reference number: JC5583; MUID:97350865  
A:Accession: JC5583  
A>Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-646 <OHA>  
A:Cross-references: GB:U96634  
A:Experimental source: thymus  
C:Comment: This protein interacts with proteins in focal complexes.  
C:Superfamily: SH3 homology; CDC24 homology  
F:13-60/Domain: SH3 homology <SH3>  
F:93-273/Domain: CDC24 homology <CD24>  
F:215-242/Domain: SH3 homology; leucine zipper #status predicted <LZP>  
F:276-279/Region: nuclear location signal  
F:408-515/Region: proline-rich  
F:508-511/Region: nuclear location signal

Query Match 12.0%; Score 123; DB 2; Length 646;  
Best Local Similarity 23.4%; Pred. No. 0.0072;  
Matches 43; Conservative 31; Mismatches 98; Indels 12; Gaps 3;  
QY 1 IAKEVSTTERTYKLEVITTSWFQSTVSKEADAMPEALAKSLIFPNFELHKEHTNFLEKEI 60  
Db 97 VLQNTILETEHEYSKELOSVLSTYLRPLQTSKLSANTSYLMGNLEEISSFOQVLVQSLE 156  
QY 61 QRLALWEGRSNAQIRDYQRIGDVMLKNIQGMKHLAAHLWKHSEALEALENGIKSSRRLE 119  
Db 157 ECTKSPEAQ-----QRVGGCFLSLMPQMRITLYLAYCANHPSAVSVL---TEHSEDLG 205  
QY 120 NCRDFELQKVCYLPNTFLRLPLRMLHYKQVLERLCKHHPPSHADRCRAALAEITE 179  
Db 206 EFMETKGASSPGILVLTGSLKSPFMRDLKPYTLLEKELERHEDYHPDRQDIQKSMATFKN 265  
QY 180 MVAQ 183  
Db 266 LSAQ 269

RESULT 9  
T42726  
guanine nucleotide release/exchange factor Ras-GRF2 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 18-Aug-2000

C:Accession: T42726  
R:Fam, N.P.; Fan, W.; Zhang, L.; Chen, H.; Moran, M.F.  
Mol. Cell. Biol. 17, 1396-1406, 1997  
A:Title: Cloning and characterization of Ras-CRF2, a novel guanine nucleotide exchange  
A:Reference number: 222245; MUID:97184464  
A:Accession: T42726  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1189 <FAM>  
A:Cross-references: EMBL:U67326; NID:gi655940; PID:gi655941; PIDN:AAC53058.1  
A:Experimental source: brain  
C:Genetics:  
A:Map position: 13  
C:Superfamily: CDC25-type guanine nucleotide exchange activator homology  
F:950-1186/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>

Query Match 11.9%; Score 122; DB 2; Length 1189;  
Best Local Similarity 23.0%; Pred.No. 0.018;  
Matches 44; Conservative 34; Mismatches 97; Indels 16; Gaps 5;

QY 9 ERTYLKDEIVTSWF-----QSTVSKEDAMPEALKSLIFPNPEPLHKFHTNFLKEIQRL 63  
| | :  
Db 255 ETEVHQIYLINGFLRPLGWAASKKPPINHDDVSISFLNSEITIMFLHEIFHQQIKARL 314  
| | :  
QY 64 ALWEGRSNAQIRDYORIGDVMKLNTQGWMKLAHLMKHSEALEENGKSSRRLENFCR 123  
| | :  
Db 315 ANWPTLVLD-----LFDILLPLMLNIYQEFVRN---HQYSLQVLAN-CKQNRFDFKLLK 364  
| | :  
QY 124 DFELOKVCI-LPLNTFFLRPLRLMHYKOVLERLCKKHPPSHADFRCDRAALAEITEMVA 182  
| | :  
Db 365 QYEANPACGEGRMLETFITPYMPQIPRYIITHELLAHTPHEHVVERKSLEFAKSKLELSR 424  
| | :  
QY 183 QLHGTMIKMEN 193  
| | :  
Db 425 VMHDEVSDTEN 435

RESULT 10  
T50454  
probable rhoL GDP-GTP exchange protein [imported] - fission yeast (Schizosaccharomyces  
C:Species: Schizosaccharomyces pombe  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 21-Jul-2000  
C:Accession: T50454  
R:McDougall, R.C.; Rajadream, M.A.; Barrell, B.G.; Simmonds, M.; Churcher, C.M.  
submitted to the EMBL Data Library, November 1999  
A:Reference number: 225030  
A:Accession: T50454  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1158 <MCD>  
A:Cross-references: EMBL:ALJ32828; PIDN:CAB60236.1; GSPDB:GNO0066; SPDB:SPAC1006.06  
A:Experimental source: strain 972h(-); cosmid c1006  
C:Genetics:  
A:Gene: SPDB:SPAC1006.06  
A:Map position: 1  
A:Introns: 835/1; 975/3  
C:Superfamily: CDC24 homology

	Query Match	11.9%	Score 121.5;	DB 2;	Length 1158;
	Best Local Similarity	23.1%;	Pred. No. 0.019;		
	Matches	45;	Conservative	36;	Mismatches 101; Indels 13; Gaps 3;
Qy	4	EVSTTERTYKDLVIITSFQSTVSKDAMPEALKS----	LIFPFEPPLHFHFNFLKEI	59	
		::::    : : :     : :   :			
Dd	454	EVIYTERDFVRDELYIRDFWTKPLSTSNVIPENNRQQFIRCVFHNMQIH-----AV	505		
Qy	60	EQRALWEGRSNAQIRDYQRIGDVMLNKIQGMKHLAHLWKHSALALENGIKSSRLE	119		
		:   :   :   :   :   :   :   :   :   :   :   :			
Dd	506	NSRLSNALNTQTLOPVVVNTIGDULVDVPKFEPPIYGANQAIAKFEFEREKSTNRFNA	565		
Qy	120	NFCRDFFE-LQKVCVYLPNTFTLLRPDLRLHMVHKQVLERLCKHPHPSHADFDRCRAALAIEIT	178		

[illegible]

Query Match	11.7%	Score 120;	DB 2;	Length 1244;
Best Local Similarity	22.1%	Pred. No. 0.029;		
Matches	44;	Conservative 33;	Mismatches 90;	Indels 32; Gaps
Qy	9	ERTYKLDVITTSWF-----QSTVSKEDAMPEALKSLIPNPFPLHKFTTNFLKETEQR	63	
Db	252	EAEVQQLHILVNNFLRPLRMAASSKKPPTHDDVSIFLNSTIMFLHQLFYQGLKARI	311	
Qy	64	ALWEGRNAQIRD-----YORIGDVMLKNTQGMKHUAAHLWKHUSEALELENGIKS	115	
Db	312	ASWPTLVLAFLDFLLPMLNIYQE---FVRNHQYSLQILAH-----CKON	353	
Qy	116	RRLNFCRDFEQLQVC-YLPLNTFLLRPLRLMHYKQVLERLCKHPPSHADFRDCRAAL	174	
Db	354	RDFDKLLQYEAKPDCERTLEFTFLTYPMFOIPRYILTLHELLAHPHEHVERNSLDYAK	413	
Qy	175	AEITEMVAQLHGHTMIKMN	193	
Db	414	SKLEELSRVMHDEVSETEN	432	

```

RESULT 12
T34428
hypothetical protein F55C7.7d - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34428
R:Du, Z.; Le, T.
submitted to the EMBL Data Library, November 1996
A:Description: The sequence of C. elegans cosmid F55C7.
A:Reference number: Z21524
A:Accession: T34428
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-552 <DUBZ>
A:Cross-references: PIDN:U080436; PIDN:AAC71111.1; GSPDB:GN00019; CESP:F55C7
A:Experimental source: strain Bristol N2; clone F55C7
C:Genetics:
A:Gene: CESP:F55C7.7d
A:Map position: 1
A:Introns: 15/2; 56/3; 329/3; 415/3; 494/2

```

Query Match 11.68; Score 119; DB 2; Length 552;

Best Local Similarity 25.7%; Pred. No. 0.013;  
Matches 43; Mismatches 32; Indels 30; Gaps 5;

QY 1 IAREVSTERTYLKDLLEVITTSFQSTVSKEDAMPEAL---KSLIFFNFELPHKFH-TNFL 56  
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 167 VLMVELTEQDYVKDLTSSVEGYIGNLNKKMD-LPADLVGDKKIIFANIVNILEFHKTNFL 225

QY 57 KETEORLALWEGSRNAQIRDYQRIGDVMLKNIOGMKHLAAHLWKHSLEALENGIKSSR 116  
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 226 KEIEKCSENTEAAGAAPVKYERRL-----HTLVITYCQNKPCKSDY 265

QY 117 RL-----ENFCROFELQKVCLPLNTFTLLRPLRLMHYKQVLRELCK 158  
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
Db 266 LLAQDDFEAFADTKAGLGHKVALCDLLIAPVRIMKYQLLKDIKL 312

RESULT 13  
T42739  
guanine nucleotide exchange factor unc-73a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 01-Dec-2000  
C:Accession: T42739; T42740; T34427; T34430; T34426  
R:Steven, R.; Kubiseski, T.J.; Zheng, H.; Kulkarni, S.; Mancillas, J.; Ruiz Morales, A.;  
Cell 92, 785-795, 1998  
A:A:Title: UNC-73 activates the Rac GTPase and is required for cell and growth cone migration  
A:Reference number: Z22256; MUID:98188103  
A:Accession: T42739  
A:A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-2488 <STE>  
A:A:Cross-references: EMBL:AF048834; NID:g2944395; PIDN:AAC12931.1; PID:g2944396  
A:Accession: T42740  
A:A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-1634,'QN',1637,'E' <SW2>  
A:A:Cross-references: EMBL:AF048835; NID:g2944397; PIDN:AAC12932.1; PID:g2944398  
R:Du, Z.; Le, T.  
submitted to the EMBL Data Library, November 1996  
A:A:Description: The sequence of C. elegans cosmid F55C7.  
A:Reference number: Z21524  
A:Accession: T34427  
A:A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-2488 <DUZ>  
A:A:Cross-references: EMBL:U80436; PIDN:AAC71108.1; GSPDB:GN00019; CESP:F55C7.7a  
A:A:Experimental source: strain Bristol N2; clone F55C7  
A:Accession: T34430  
A:A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-1634,'QN',1637,'E' <DUZ>  
A:A:Cross-references: EMBL:U80436; PIDN:AAC71109.1; GSPDB:GN00019; CESP:F55C7.7b  
A:A:Experimental source: strain Bristol N2; clone F55C7  
A:Accession: T34426  
A:A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 'MVIKCFM',1568,'LRK',1572-2488 <DUZ>  
A:A:Cross-references: EMBL:U80436; PIDN:AAC71110.1; GSPDB:GN00019; CESP:F55C7.7c  
A:A:Experimental source: strain Bristol N2; clone F55C7  
C:Genetics:  
A:Gene: unc-73; F55C7.7a; unc-73; F55C7.7b; CESP:F55C7.7c  
A:Map position: 1  
A:Introns: 7/3; 64/2; 95/3; 230/3; 265/2; 335/3; 360/3; 488/3; 569/1; 695/1; 1028/2; 106  
19/3; 2262/2; 2331/1; 2351/3; 2396/1; 2431/1

```

Query Match      11.8%; Score 119; DB 2; Length 2488;
Best Local Similarity 25.7%; Pred. No. 0.08;
Matches 43; Conservative 30; Mismatches 62; Indels 30; Gaps 5;
QY 1 IAKEVSTERTYLDKLVLSVITFSWFOSTVSKEDAMPEAL---KSLFFNPPEPLHKHF-TNFL 56
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 1813 VLMLVETQDYVVKDLRSVVEGYVIGLNKMD-LEADLVGKKDKIIFATVINILEFHKHTNFL 1871

```

[illegible]

```

RESULT      15
TVHUDB
transforming protein dbl precursor - human
C:Species: Homo sapiens (man)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 02-Sep-1997
C:Accession: A30040
R:Ron, D.; Tronick, S.R.; Aaronson, S.A.; Eva, A.
EMBO J. 7, 2465-2473, 1988
A:Title: Molecular cloning and characterization of the human dbl proto-oncogene: evidence for a novel protein-coding region
A:Reference number: A30040; MUID:89052660
A:Accession: A30040
A:Molecule type: mRNA
A:Residues: 1-925 <RON>
A:Cross-references: EMBL:X12556
C:Genetics:
A:Gene: GDB:MCF2
A:Cross-references: GDB:I20168; OMIM:311030
A:Map position: Xq26.3-Xq27.1
C:Superfamily: dbl transforming protein; CDC24 homology; pleckstrin repeat homology
C:Keywords: oncogene; phosphoprotein; transforming protein

```

F;495-675/Domain: CDC24 homology <CD24>  
F;498-925/Product: transforming protein dbl #status predicted <DBL>  
F;703-807/Domain: pleckstrin repeat homology <PLK>

```
Query Match      11.5%; Score 118; DB 1; Length 925;
Best Local Similarity 21.0%; Pred. No. 0.03;
Matches 44; Conservative 45; Mismatches 67; Indels 54; Gaps 9;

QY 1 IAKEVSTTERTYTKDLEVITSNFQSTVSKE---DAMPEAL---KSLIFPNFELHKFHTN 54
   : ||| |::| : :: : | || | | : | : | : | : | : | : | : | :
Db 499 VLNELIQTERVYVRELYTVLLGYRAEMDNPEFDLMPPLRNKKDILFGNMAEIEFHND 558
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 55 FLKEIEQRLALWEGRSNAQIRDYRIGDVMLKNIQGMKHLAAH-----LW-KHSEA 104
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 559 IF-----LSSLENCAP-----ERYGPGFLERKDDFQMYAKYCNKPRSEIWRKYSEC 608
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 105 LEALENGIKSSRRLENFCRDFELQKVCYLPLNTFLRLPLRLMHYKQVLERLCKKHPPSH 164
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 609 AFFEQCQRLKHRLR-----LDSYLLKPVQRITKYQLLKLKLYS----- 649
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 165 ADPRDC-----RAALAEITEMVAQLHGTM 188
   : || : || : :: : : : |
Db 650 ---KCEGSALLKKALDAMDLLKSYNDSM 676
   : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
```

Search completed: December 6, 2001, 08:52:39  
Job time: 371 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:53:08 ; Search time 28.69 Seconds  
(without alignments)  
247.925 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_544\_737

Perfect score: 1023

Sequence: 1 IAKEVSTTERTYKDLVIT.....AEITEMVAQLHGTMIKMF 194

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	154	15.1	960	1 FGD1_MOUSE	P52734 mus musculus
2	154	15.1	961	1 FGD1_HUMAN	P98174 homo sapien
3	138	13.5	3038	1 TRIO_HUMAN	O75962 homo sapien
4	131	12.8	646	1 Y142_HUMAN	Q14155 homo sapien
5	129.5	12.7	596	1 LFC_MOUSE	Q60875 mus musculus
6	129.5	12.3	893	1 GEFH_HUMAN	Q92974 homo sapien
7	120	11.7	1244	1 GNRP_RAT	P28818 rattus norv
8	118	11.5	925	1 DBL_HUMAN	P10911 homo sapien
9	117.5	11.5	738	1 ECT2_MOUSE	Q07139 mus musculus
10	117.5	11.5	2044	1 SIF2_DROME	P91620 drosophila
11	117.5	11.5	2064	1 SIF1_DROME	P91621 drosophila
12	116	11.3	872	1 SCD1_SCHPO	P40995 schizosacch
13	116	11.3	1262	1 GNRP_MOUSE	P27671 mus musculus
14	114	11.1	854	1 CC24_YEAST	P11433 saccharomyc
15	111	10.9	937	1 DBS_RAT	Q63406 rattus norv
16	111	10.9	1275	1 GNRP_HUMAN	Q13972 homo sapien
17	109.5	10.7	1108	1 DBS_HUMAN	O15068 homo sapien
18	108	10.6	845	1 VAV_MOUSE	P27870 mus musculus
19	106	10.4	845	1 VAV_HUMAN	P15498 homo sapien
20	105.5	10.3	1693	1 RIP2_MOUSE	P97433 mus musculus
21	105	10.3	1271	1 BCR_HUMAN	P11274 homo sapien
22	104	10.2	1591	1 TIAM_HUMAN	Q13009 homo sapien
23	104	10.2	1591	1 TIAM_MOUSE	Q06610 mus musculus
24	103.5	10.1	847	1 VAV3_HUMAN	Q9UKW4 homo sapien
25	103	10.1	843	1 VAV_RAT	P54100 rattus norv
26	103	10.1	878	1 VAV2_HUMAN	P52735 homo sapien
27	102	10.0	868	1 VAV2_MOUSE	Q60992 mus musculus
28	101	9.9	519	1 TIM_HUMAN	Q12774 homo sapien
29	99	9.7	1663	1 HAPP_HUMAN	O60229 homo sapien
30	98	9.6	1919	1 HAPP_RAT	P97924 rattus norv
31	96.5	9.4	2137	1 SPCB_HUMAN	P11277 homo sapien
32	95	9.3	1149	1 DBS_MOUSE	Q64096 mus musculus
33	94.5	9.2	847	1 VAV3_MOUSE	Q9P0C8 mus musculus

#### RESULT 1 FGD1\_MOUSE

ID	FGD1_MOUSE	STANDARD;	PRT;	960 AA.
AC	P52734;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)			
DE	(FACIOGENITAL DYSPLASIA PROTEIN HOMOLOG).			
GN	FGD1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=96081343; PubMed=8535076;			
RA	Pasteris N.G., de Gouyon B., Cadle A.B., Campbell K., Herman G.E.,			
RA	Gorski J.L.			
RT	"Cloning and regional localization of the mouse facio-genital			
RT	dysplasia (Fgdl) gene."			
RL	Mamm. Genome 6:658-661(1995).			
CC	-!- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS			
CC	BY EXCHANGING BOUND GDP FOR FREE GTP.			
CC	-!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).			
CC	-!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).			
CC	-!- SIMILARITY: CONTAINS 2 PH DOMAINS.			
CC	-!- SIMILARITY: CONTAINS 1 FYVE-TYPE ZINC FINGER.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
DR	EMBL; U22325; AAA96001.1; -			
DR	MGD; MGI:104566; Fgdl.			
DR	InterPro; IPR001849; PH.			
DR	InterPro; IPR000219; RhoGEF.			
DR	InterPro; IPR000822; Znf-C2H2.			
DR	InterPro; IPR000306; Znf_FYVE.			
DR	Pfam; PF01363; FYVE; 1.			
DR	Pfam; PF00169; PH; 2.			
DR	Pfam; PF00621; RhoGEF; 1.			
DR	SMART; SM00064; FYVE; 1.			
DR	SMART; SM00233; PH; 2.			
DR	SMART; SM00325; RhoGEF; 1.			
DR	PROSITE; PS50178; ZF_FYVE; 1.			
DR	PROSITE; PS50003; PH_DOMAIN; 2.			
KW	Guanine-nucleotide releasing factor; zinc-finger.			
FT	DOMAIN 7 330			
FT	PRO-RICH.			
FT	DOMAIN 171 179			
FT	SH3-BINDING (POTENTIAL).			
FT	DOMAIN 179 187			
FT	SH3-BINDING (POTENTIAL).			
FT	DOMAIN 569 688			
FT	PH 1.			

#### ALIGNMENTS

```
FT ZN_FING 729 789 FIVE-TYPE.
FT DOMAIN 820 920 PH 2.
SQ SEQUENCE 960 AA; 106477 MW; 41C1B84DE490FC51 CRC64;

Query Match 15.1%; Score 154; DB 1; Length 960;
Best Local Similarity 23.7%; Pred. No. 1.le-05;
Matches 47; Conservative 41; Mismatches 96; Indels 14; Gaps 4;

QY 1 IAEKSTERTYKLDLEVTWFSQSTVSKE----DAMPEALKSLIFPNFEPLHKFTNF- 55
DB 376 IANELLQTEKAYVSRHLHLLDQVFCARLLEARNRSPADVVGIFSNICSICYCFHQQFL 435
QY 56 LKEIEQRLALWEGRSNAQIRDYQRIQDVMLKNIQGMKHLAAHLWKHSEALEALENGIKSS 115
DB 436 LPELEKRMEDW-----RYPRIGDILQKLAPFLKMYGVYKVFNDRAVELVNTWTWTS 487
QY 116 RLLENFCRDFELQKVC-YLPLNTFLRLPLRLMHYQVLERLCKKHPPSHADFRDRAAL 174
DB 488 TQFKVIHEVQKEACGNLTQHHMLEPVQIRPYVELLLKDYLLKLPHGSPDSKDAQSL 547
QY 175 AEITENVAQLHGTMIKME 192
DB 548 ELIATAEAHSNAAIRKME 565

RESULT 2
ID FGD1_HUMAN STANDARD; PRT; 961 AA.
AC P98174;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)
DE (FACIOGENITAL DYSPLASIA PROTEIN).
GN FGD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Craniofacial;
RX MEDLINE=95042764; PubMed=7954831;
RA Pastoris N.G., Cadle A., Logie L.J., Porteous M.E.M., Schwartz C.E.,
RA Stevenson R.E., Glover T.W., Wilroy R.S., Gorski J.L.;
RT "Isolation and characterization of the facio-genital dysplasia
RT (Aarskog-Scott syndrome) gene: a putative Rho/Rac guanine nucleotide
RT exchange factor.";
RL Cell 79:669-678(1994).
CC -!- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS
CC BY EXCHANGING BOUND GTP FOR FREE GTP.
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: FETAL HEART, BRAIN, LUNG, KIDNEY, PLACENTA,
CC AND LESS IN LIVER; ADULT HEART, BRAIN, LUNG, SKELETAL MUSCLE, AND
CC LESS IN PANCREAS AND LIVER.
CC -!- DISEASE: FACIOGENITAL DYSPLASIA (AARSKOG-SCOTT SYNDROME) IS A RARE
CC MULTISYSTEMIC DISORDER CHARACTERIZED BY DISPROPORTIONATELY SHORT
CC STATURE, AND BY FACIAL, SKELETAL, AND UROGENITAL ANOMALIES.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 FIVE-TYPE ZINC FINGER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U11690; AAA57004.1; -
CC MIM; 305400; -
CC
CC STRUCTURE BY NMR OF 1227-1407, AND MUTAGENESIS.
```

```
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR000822; Znf-C2H2.
DR InterPro: IPR000306; Znf_FYVE.
DR Pfam: PF01363; FYVE; 1.
DR Pfam: PF00169; PH; 2.
DR Pfam: PF00621; RhoGEF; 1.
DR SMART: SM00064; FYVE; 1.
DR SMART: SM00233; PH; 2.
DR SMART: SM00325; RhoGEF; 1.
DR PROSITE: PS50178; ZF_FYVE; 1.
DR PROSITE: PS50003; PH_DOMAIN; 2.
KW Guanine-nucleotide releasing factor; zinc-finger.
FT DOMAIN 7 330 PRO-RICH.
FT DOMAIN 171 179 SH3-BINDING (POTENTIAL).
FT DOMAIN 179 187 SH3-BINDING (POTENTIAL).
FT DOMAIN 590 689 PH 1.
FT ZN_FING 730 790 FYVE-TYPE.
FT DOMAIN 821 921 PH 2.
SQ SEQUENCE 961 AA; 106668 MW; CF722598853A685A CRC64;

Query Match 15.1%; Score 154; DB 1; Length 961;
Best Local Similarity 23.7%; Pred. No. 1.le-05;
Matches 47; Conservative 41; Mismatches 96; Indels 14; Gaps 4;

QY 1 IAEKSTERTYKLDLEVTWFSQSTVSKE----DAMPEALKSLIFPNFEPLHKFTNF- 55
DB 377 IANELLQTEKAYVSRHLHLLDQVFCARLLEARNRSPADVVGIFSNICSICYCFHQQFL 436
QY 56 LKEIEQRLALWEGRSNAQIRDYQRIQDVMLKNIQGMKHLAAHLWKHSEALEALENGIKSS 115
DB 437 LPELEKRMEDW-----RYPRIGDILQKLAPFLKMYGVYKVFNDRAVELVNTWTWTS 488
QY 116 RLLENFCRDFELQKVC-YLPLNTFLRLPLRLMHYQVLERLCKKHPPSHADFRDRAAL 174
DB 489 TQFKVIHEVQKEACGNLTQHHMLEPVQIRPYVELLLKDYLLKLPHGSPDSKDAQSL 548
QY 175 AEITENVAQLHGTMIKME 192
DB 549 ELIATAEAHSNAAIRKME 566

RESULT 3
ID TRIO_HUMAN STANDARD; PRT; 3038 AA.
AC Q75962; Q13458;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE TRIPLE FUNCTIONAL DOMAIN PROTEIN (PTPRF INTERACTING PROTEIN).
GN TRIO.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Fibroblast;
RX MEDLINE=96224308; PubMed=8643598;
RA Debat A., Serra-Pages C., Seipel K., O'Brien S., Tang M., Park S.-H.,
RA Streuli M.;
RT "The multidomain protein Trio binds the LAR transmembrane tyrosine
RT phosphatase, contains a protein kinase domain, and has separate rac-
RT specific and rho-specific guanine nucleotide exchange factor
RL Proc. Natl. Acad. Sci. U.S.A. 93:5466-5471(1996).
[2]
RN
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Streuli M.;
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
[3]
RN
RP STRUCTURE BY NMR OF 1227-1407, AND MUTAGENESIS.
```



```
Db 2012 IVSEYIDTFPDLKQRLGHRQLTDLLIKPVQRMKYQLLKDLKFLKYSKASLDTSER 2071
QY 173 AL 174
   |
Db 2072 AV 2073
   |

RESULT 4
ID Y142_HUMAN STANDARD; PRT; 646 AA.
AC Q14155;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA0142.
GN KIAA0142.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone marrow;
RX MEDLINE=96127530; PubMed=8590280;
RA Nagase T., Seki N., Tanaka A., Ishikawa K.-I., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. IV.
RT The coding sequences of 40 new genes (KIAA0121-KIAA0160) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 2:167-174(1995).
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63476; BAA09763.1; -.
DR HSSP; P19174; IHSQ.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
KW Hypothetical protein, SH3 domain.
FT DOMAIN 6 65
FT DOMAIN 295 400
FT SEQUENCE 646 AA; 73140 MW; B5B5A83F0EBC28D2 CRC64;

Query Match 12.8%; Score 131; DB 1; Length 646;
Best Local Similarity 25.1%; Pred. No. 0.00064;
Matches 49; Conservative 31; Mismatches 81; Indels 34; Gaps 5;

QY 1 IAKEVSTERTYTKDLKLEVTISWFSQSVSKEDAMPEALKSLIFNFPFLKFTNFKETE 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 97 VLQNILETENEYSKELOTVLSYLRPLQTSKLSANISYLMGNLEICSFQOMLVQSLE 156
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 QRALMEGRSNAQIRDYQRIGDVMLKNIQMKHL-----AAH-----LWKHSEAL-EAL 108
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 157 ECTKLPEAQ-----QRVGGCFLMLPQMKTLTYLCANHPNSAVNVLTEHSEELGEP 208
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 109 ENGIKSSRLNEFCRDFELQKVCYLPPLNTFLRLPLRLMHYKOVLRCLCKHHPPSHADFR 168
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 209 ETKGASSPGI-----LVLTTCLSKPFMRDLKDYPTLLKELERHMHEDYHTDRQ 254
QY 169 DCRAALAEITEMVAQ 183
   | : : | |
Db 255 DIQKSMAAFKNLSAQ 269
   | : : | |

RESULT 5
ID LFC_MOUSE STANDARD; PRT; 596 AA.
AC Q60875; O09115;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE LYMPHOID BLAST CRISIS-LIKE 1 (LBC'S FIRST COUSIN) (ONCOGENE LFC)
DE (RHOBIN).
GN LBC1L OR LFC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Myeloid;
RX MEDLINE=95355462; PubMed=7629163;
RA Whitehead I., Kirk H., Tognon C., Trigo-Gonzalez G., Kay R.;
RT "Expression cloning of lfc, a novel oncogene with structural
RT similarities to guanine nucleotide exchange factors and to the
RT regulatory region of protein kinase C.";
RL J. Biol. Chem. 270:18388-18395(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C;
RA Olofsson B.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS, WITH THE EXCEPTION OF LIVER
CC TISSUE. LEVELS ARE HIGH IN HEMOPOIETIC TISSUES (THYMUS, SPLEEN,
CC BONE MARROW) AS WELL AS IN KIDNEY AND LUNG.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.
CC -!- SIMILARITY: TO HUMAN NUCLEOTIDE EXCHANGE PROTEIN LBC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U28495; AAC52234.1; -.
DR EMBL; X95761; CAA65067.1; -.
DR MGD; MGI:103284; Lbcll.
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
DR Pfam; PF00130; DAG_PE-bind; 1.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR SMART; SM00109; CL; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE; PS00081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE; PS50003; PH DOMAIN; 1.
KW Guanine-nucleotide releasing factor; Proto-oncogene;
KW Phorbol-ester binding; Zinc.
FT DOMAIN 40 86
FT DOMAIN 240 511
FT DOMAIN 487 573
FT DOMAIN 493 496
FT POLY-LEU.
```

```
FT CONFLICT 1 21 MSRIEQLTRARIDRSKEQATK -> MSGNRRPSPRRGQ
FT CONFLICT 156 156 A -> V (IN REF. 2).
FT CONFLICT 574 596 MISSING (IN REF. 1).
SQ SEQUENCE 596 AA; 68584 MW; E168508BDC7C6E13 CRC64;

Query Match 12.7%; Score 129.5; DB 1; Length 596;
Best Local Similarity 21.4%; Pred. No. 0.00078;
Matches 44; Conservative 50; Mismatches 85; Indels 27; Gaps 7;

QY 4 EVSTTERTYKLEVITTSWFQSTVSKEDAM-PEALKSLIFPNEPLKHFTHFLKEI--E 60
Db 243 ELTQTELHHVRTLKIMTRFLFTGMLEELHLEPGVVGQGL-FPCVDELSDIHTFLNQLLER 301

QY 61 QRLAWEGRSNAQIRDYQIRIGDVMKNIQ-----MKHLAAHLWKHSEALENGIKS 114
Db 302 RQALCPGSTRNFV--IHLRGDLLISQFSGSNAEQMRKTYSEFCSRHTALKLYKELYAR 359

QY 115 SRLENFCRDFELQKV-----CYLPNTFLRLPLRLMHYKQVLERCKLKHPPSHAD 166
Db 360 DRFOQFIRKTRSAVKLRHGVQEC-----ILLVTRITKYPVLNIRLQNSHGVEE 412

QY 167 FRDCRAALAEITEMVAQLHGTMIKME 192
Db 413 YQDLASALGLVKELLSNVDDVHELE 438

RESULT 6
GEPH_HUMAN
ID GEPH_HUMAN STANDARD; PRT; 893 AA.
AC Q92974; Q15079; 075142;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GEF-H1 PROTEIN (PROLIFERATING CELL NUCLEOLAR ANTIGEN P40).
GN LFP40 OR KIA0651.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99074271; PubMed=9857026;
RA Ren Y., Li R., Zheng Y., Busch H.;
RT "Cloning and characterization of GEF-H1, a microtubule-associated
RT guanine nucleotide exchange factor for Rac and Rho GTPases.";
RL J. Biol. Chem. 273:34954-34960(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98403880; PubMed=9734811;
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. X.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 5:169-176(1998).
RN [3]
RP REVISIONS.
RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 630-893 FROM N.A.
RX MEDLINE=89168219; PubMed=2466560;
RA Reddy A.B., Chatterjee A., Rothblum L.I., Black A., Busch H.;
RT "Isolation and characterization of complementary DNA to proliferating
RT cell nucleolar antigen p40.";
RL Cancer Res. 49:1763-1767(1989).
CC -!- FUNCTION: STIMULATES GUANINE NUCLEOTIDE EXCHANGE OF RAC AND RHO
CC BUT IS INACTIVE TOWARD CDC42, TC10, OR RAS. BINDS TO RAC AND RHO
CC PROTEINS IN BOTH THE GDP-AND GUANOSINE 5'-3-O-(THIO)TRIPHOSPHATE-
```

```
CC BOUND STATES WITHOUT DETECTABLE AFFINITY FOR CDC42 OR RAS. MAY
CC HAVE A DIRECT ROLE IN ACTIVATION OF RAC AND/OR RHO AND IN BRINGING
CC THE ACTIVATED GTPASE TO SPECIFIC TARGET SITES SUCH AS
CC MICROTUBULES.
CC -!- SUBCELLULAR LOCATION: COLOCALIZES WITH MICROTUBULES THROUGH THE
CC CARBOXYL-TERMINAL COILED-COIL DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.
CC -!- SIMILARITY: COULD BE THE ORTHOLOG OF MOUSE LYMPHOID BLAST CRISIS-
CC LIKE 1 (LFC ONCOGENE).
CC -!- CAUTION: THE SEQUENCE SHOWN HERE COMES FROM THE FIGURE OF REF.1,
CC APPARENTLY THE SUBMITTED SEQUENCE IS WRONG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U72206; AAC97383.1; ALT_SEQ.
CC EMBL; AB014551; BAA31626.2; -.
CC EMBL; X15610; CAA33634.1; -.
CC InterPro; IPR002219; DAG_PE-bind.
CC InterPro; IPR001849; PH.
CC InterPro; IPR000219; RhoGEF.
CC Pfam; PF00130; DAG_PE-bind; 1.
CC Pfam; PF00169; PH; 1.
CC Pfam; PF00621; RhoGEF; 1.
CC SMART; SM00109; C1; 1.
CC SMART; SM00233; PH; 1.
CC SMART; SM00325; RhoGEF; 1.
CC PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
CC PROSITE; PS50081; DAG_PE_BIND_DOM_2; 1.
CC PROSITE; PS50003; PH_DOMAIN; 1.
CC Guanine-nucleotide releasing factor; Phorbol-ester binding; Zinc;
CC Coiled coil.
CC DOMAIN 40 86 PHORBOL-ESTER AND DAG BINDING.
CC DOMAIN 238 509 DH.
CC DOMAIN 485 571 PH.
CC DOMAIN 589 610 COILED COIL (POTENTIAL).
CC DOMAIN 797 868 COILED COIL (POTENTIAL).
CC CONFLICT 1 21 MSRIEQLTRARIDRSRELASK -> LGSAHAGMCCCGCCP
CC CONFLICT 193 193 LLAHLEQGSLLER (IN REF. 2).
CC CONFLICT 579 579 E -> EA (IN REF. 2).
CC CONFLICT 867 867 S -> P (IN REF. 2).
CC CONFLICT 885 885 Q -> P (IN REF. 2).
CC CONFLICT SSPQAMPCT -> SLPAGDALYLSFNPPQPSRGTRDLPLV
CC (IN REF. 2).
CC SEQUENCE 893 AA; 101173 MW; 80AC8FA7F762E9C3 CRC64;

Query Match 12.3%; Score 125.5; DB 1; Length 893;
Best Local Similarity 20.5%; Pred. No. 0.0027;
Matches 41; Conservative 51; Mismatches 93; Indels 15; Gaps 5;

QY 4 EVSTTERTYKLEVITTSWFQSTVSKEDAMPEALKSLIFPNEPLKHFTHFLKEI--EQ 61
Db 241 ELTQTELHHVRTLKIMTRFLFTGMLEELHLEPGVVGQGLFPCVDELSDIHTFLNQLLER 300

QY 62 RIALWEGRSNAQIRDYQIRIGDVMKNIQ-----MKHLAAHLWKHSEALENGIKS 115
Db 301 RQALCPGSTRNFV--IHLRGDLLISQFSGPSAEQMRKTYSEFCSRHSKALKLYKELYARD 358

QY 116 RRLNEFCRDFELQKVCLP---LNTFLRLPLRLMHYKQVLERCKLKHPPSHADFRDCA 172
Db 359 KRQEQFIR--KVTRPAVLKRHGVQECILLVTRITKYPVLNIRLQNSHGIEERDLTT 416

QY 173 ALAEITEMVAQLHGTMIKME 192
Db 413 YQDLASALGLVKELLSNVDDVHELE 438
```

```
Db 417 ALGLVKELLSNVDEGIYQLE 436
RESULT 7
GNRP_RAT STANDARD; PRT; 1244 AA.
AC P28818;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GUANINE NUCLEOTIDE RELEASING PROTEIN (GNRP) (P140 RAS-GRF).
GN RASGRF1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92350260; PubMed=1379346;
RA Shou C., Farusworth C.L., Neel B.G., Feig L.A.;
RT "Molecular cloning of cDNAs encoding a guanine-nucleotide-releasing
factor for Ras p21";
RL Nature 358:351-354(1992).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X67241; CAA47666.1; -
DR PIR; S29083; S29083.
DR InterPro; IPR001331; GDS_CDC24.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000651; RasGEFN.
DR InterPro; IPR001895; RasGRF_CDC25.
DR InterPro; IPR000219; RhoGEF.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00169; PH; 2.
DR Pfam; PF00617; RasGEF; 1.
DR Pfam; PF00618; RasGEFN; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEFN; 2.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00720; GDS_CDC25; 1.
DR PROSITE; PS00741; GDS_CDC24; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
DR PROSITE; PS50096; IQ; 1.
KW Guanine-nucleotide releasing factor.
FT DOMAIN 22 129 PH 1.
FT DOMAIN 204 229 IQ.
FT DOMAIN 244 455 DH.
FT DOMAIN 456 582 PH 2.
FT DOMAIN 1007 1241 RASGEF.
SQ SEQUENCE 1244 AA; 142666 MW; 4B647879E842AF6B CRC64;

Query Match 11.7%; Score 120; DB 1; Length 1244;
Best Local Similarity 22.1%; Pred. No. 0.012;
Matches 44; Conservative 33; Mismatches 90; Indels 32; Gaps 5;
```

```
QY 9 ERTYLDKLEVTISWF-----QSTVSKEDAMPEALKSLIFPFPLHKHFTNFKLEIEQRL 63
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 252 EAEYVQQLHILVNNFLRPLRMAASSKPPITHDDVSSIFLNSSETIMFLHQIFYQGLKARI 311
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 64 ALWEGSRNAQIRD-----YQIRGDVMLKNIQGMKHLAAHLWKHSEALEALENGIKSS 115
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 312 ASWPTLVLDLFDILLPLMLNIQE---FVRNHOYSLQILAH-----CKQN 353
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 116 RRLNECFRQFELQKVC-YLPLNFTFLRLHRLMHYQVLERLCKHHPPSHADFRDCRAAL 174
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 354 RQFDKLLKQYEAKPDCEERTLETFLYPPQIPRYILTLHELLAHTPHEVERNSLDYAK 413
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 175 AEITEMVAQLHGTMIKWN 193
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 414 SKLEELSRVMHDEVSETN 432

RESULT 8
DBL_HUMAN STANDARD; PRT; 925 AA.
ID DBL_HUMAN
AC P10911; P14919;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTO-ONCOGENE DBL PRECURSOR [CONTAINS: MCF2].
GN DBL OR MCF2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=89052660; PubMed=3056717;
RA Ron D., Tronick S.R., Aaronson S.A., Eva A.;
RT "Molecular cloning and characterization of the human dbl proto-
oncogene: evidence that its overexpression is sufficient to transform
NIH/3T3 cells.";
RL EMBO J. 7:2465-2473(1988).
RN [2]
RP REVISIONS.
RA Ron D.;
RL Submitted (JUN-1989) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 498-925 FROM N.A.
RX MEDLINE=88176886; PubMed=3281159;
RA Eva A., Vecchio G., Rao C.D., Tronick S.R., Aaronson S.A.;
RT "The predicted DBL oncogene product defines a distinct class of
transforming proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:2061-2065(1988).
RN [4]
RP SEQUENCE OF 398-925 FROM N.A.
RX MEDLINE=94167115; PubMed=2577874;
RA Noguchi T., Galland F., Batzo M., Mattei M.-G., Birnbaum D.;
RT "Activation of a mcf.2 oncogene by deletion of amino-terminal coding
sequences.";
RN [5]
RP Oncogene 3:709-715(1988).
RX DBL-HOMOLOGY DOMAIN, AND MUTAGENESIS.
RX MEDLINE=91291759; PubMed=2065022;
RA Ron D., Zannini M., Lewis M., Wickner R.B., Hunt L.T., Graziani G.,
RA Tronick S.R., Aaronson S.A., Eva A.;
RT "A region of proto-dbl essential for its transforming activity shows
sequence similarity to a yeast cell cycle gene, CDC24, and the human
breakpoint cluster gene, bcr.";
RL New Biol. 3:372-379(1991).
RN [6]
RP CHARACTERIZATION OF DBL DOMAIN.
RX MEDLINE=94103281; PubMed=8276860;
RA Hart M.J., Eva A., Zangrilli D., Aaronson S.A., Evans T.,
RA Cerione R.A., Zheng Y.;
RT "Cellular transformation and guanine nucleotide exchange activity are
catalyzed by a common domain on the dbl oncogene product.";
```

J. Biol. Chem. 269:62-65(1994).

-!- FUNCTION: PROMOTES THE EXCHANGE OF G25K-BOUND GDP BY GTP. ACTS AS A GUANINE NUCLEOTIDE EXCHANGE FACTOR FOR CDC42 AND RHOA, BUT NOT FOR RAC1 AND TC10. CAN BE ACTIVATED BY TRUNCATION OF THE N-TERMINUS.

-!- SUBCELLULAR LOCATION: CYTOPLASMIC.

-!- TISSUE SPECIFICITY: EXPRESSED IN FETAL BRAIN AND ADRENAL GLANDS; IN THE ADULT, EXPRESSION IS LIMITED TO TESTES, OVARIES AND BRAIN.

-!- DOMAIN: THE DH DOMAIN IS ESSENTIAL FOR TRANSFORMING ACTIVITY AND DIRECTLY CATALYZES GDP-GTP EXCHANGE ACTIVITY.

-!- DISEASE: MCF.2 AND DBL REPRESENT TWO ACTIVATED VERSIONS OF THE SAME PROTO-ONCOGENE.

-!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).

-!- SIMILARITY: CONTAINS 1 PH DOMAIN.

This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; X12556; CAA31069.1; -  
DR EMBL; J03639; AAS2172.1; ALT\_INIT.  
DR EMBL; X13230; CAA31617.1; ALT\_SEQ.  
DR PIR; A30040; TVH0DB.  
DR PIR; A28051; TVH0DB.  
DR MIM; 311030; -  
DR InterPro; IPR000947; CBFA\_NFYB.  
DR InterPro; IPR001251; CRAL\_FR2A.  
DR InterPro; IPR001331; GDS\_CDC24.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR000219; RhoGEF.  
DR InterPro; IPR002017; Spectrin.  
DR Pfam; PF00169; PH; 1.  
DR Pfam; PF00621; RhoGEF; 1.  
DR PRINTS; PR00615; CCAATSUBUNTA.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00325; RhoGEF; 1.  
DR SMART; SM00150; SPEC; 1.  
DR PROSITE; PS00741; GDS\_CDC24; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
KW Guanine-nucleotide releasing factor; Proto-oncogene; Phosphorylation; Transforming protein.  
FT CHAIN 398 925 MCF2 TRANSFORMING PROTEIN.  
FT CHAIN 498 925 DBL TRANSFORMING PROTEIN.  
FT DOMAIN 498 686 DH.  
FT DOMAIN 687 809 PH.  
FT MOD\_RES 742 742 PHOSPHORYLATION.  
FT MUTAGEN 640 646 CAPABILITY REDUCED; NO STIMULATION OF GDP DISSOCIATION.  
FT CONFLICT 634 634 R -> Q (IN REF. 4).  
FT CONFLICT 886 886 A -> V (IN REF. 3).  
FT SEQUENCE 925 AA; 107657 MW; 0BCCB303276D19C9 CRC64;  
SQ  
Query Match 11.5%; Score 118; DB 1; Length 925;  
Best Local Similarity 21.0%; Pred. No. 0.013;  
Matches 44; Conservative 45; Mismatches 67; Indels 54; Gaps 9;  
QY 1 IAKEVSTERTYKLDLEVTISWFQSTVSKE---DAMPEAL---KSLFPNFEPLHKEFTN 54  
Db 499 VLNELIQTERVYVRELYTVLLGYSRAEMDNPEMFDLPLLRNKKDKILFGNMAEYEFHND 558  
QY 55 FLKEISQRIALWEGRNAQIRVQTDVWLKNIOGKKHLAAH-----LW-KHSEA 104  
Db 559 IF-----LSSLENCAHAP-----ERVGPCFLERKDDFQMYAKYCKQNKPRSETIWRKYSEC 608  
QY 105 LEALENGIKSSRRLENFCRDFELQKVCYLPNTFLRLPLRLHMYKOVLERLCKKHPPSH 164  
Db 609 AFFQECQRLKHLR-----LDSYLLKPVQIRITYQLLLKELLKYS----- 649

QY 165 ADRPDC-----RAALAEITEMVAQLHGTM 188  
Db 650 ---KDCGSALLKALDAMDLLKSYNDMS 676

RESULT 9  
ECT2\_MOUSE  
ID ECT2\_MOUSE STANDARD; PRT; 738 AA.  
AC Q07139;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE ECT2 PROTEIN (ECT2 ONCOGENE).  
GN ECT2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93218723; PubMed=8464478;  
RA Miki T., Smith C.L., Long J.E., Eva A., Fleming T.P.;  
RT "Oncogene ect2 is related to regulators of small GTP-binding proteins";  
RT Nature 362:462-465(1993).  
RL Nature 362:462-465(1993).  
RN [2]  
RP ERRATUM.  
RA Miki T., Smith C.L., Long J.E., Eva A., Fleming T.P.;  
RL Nature 364:737-737(1993).  
CC -!- FUNCTION: TRANSFORMING PROTEIN THAT BINDS HIGHLY SPECIFICALLY TO RHOA, RHOE AND RAC PROTEINS, BUT DOES NOT APPEAR TO CATALYZE GUANINE NUCLEOTIDE EXCHANGE. TRUNCATION OF THE N-TERMINUS PUTS THE PROTEIN IN AN ACTIVATED STATE.  
CC -!- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS. ALSO DETECTABLE IN KIDNEY, LIVER AND SPLEEN.  
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; L11316; AAA37536.1; -  
DR MGD; MGI:95281; Ect2.  
DR InterPro; IPR001357; BRCT.  
DR InterPro; IPR001331; GDS\_CDC24.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR000219; RhoGEF.  
DR Pfam; PF00533; BRCT; 2.  
DR Pfam; PF00621; RhoGEF; 1.  
DR SMART; SM00292; BRCT; 2.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00325; RhoGEF; 1.  
DR PROSITE; PS0172; BRCT; 2.  
DR PROSITE; PS00741; GDS\_CDC24; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
KW Guanine-nucleotide releasing factor; Proto-oncogene.  
FT DOMAIN 280 499 DH.  
FT DOMAIN 500 619 PH.  
SQ SEQUENCE 738 AA; 83685 MW; 3D270CBF1570C681 CRC64;

Query Match 11.5%; Score 117.5; DB 1; Length 738;  
Best Local Similarity 22.3%; Pred. No. 0.011;  
Matches 50; Conservative 30; Mismatches 79; Indels 65; Gaps 8;  
QY 1 IAKEVSTERTYKLDLEVTISWFQSTVSKE-----AMPEALKSLFPNFEPLHKEFTN 54  
Db 1 IAKEVSTERTYKLDLEVTISWFQSTVSKE-----AMPEALKSLFPNFEPLHKEFTN 54

Db 281 VAKELYTESNYVILATIIQLFOVPLEEGRGGPILAPEIKT-IFGSIPIDFVHMK 339  
QY 55 FLKEIEQRLALW-EGRNAQIRDYQIGDVMLKNIQGMKHLAAHLWKHSEALELENGIK 113  
Db 340 IKDDLEDLIANWDESR-----IGDFELKAKDL-----VK 370  
QY 114 SSRLENFCRDFELQVKY-----CYPLN-----TFLLRPLRLHLM 148  
Db 371 TYPEPVNF---FEMSEKMIKCEKQKPRFHAFLKINAKPEGCRQSLVELLIRPVQLRPS 427  
QY 149 YKQVLERLCKHPPSHADFRCRAALAEITEMVAQLHGTMKWE 192  
Db 428 VALLNDLKKHTADENPKSTLEKAIGSLKEVMTNIEDRRKTE 471  
  
RESULT 10  
SIF2\_DROME STANDARD; PRT; 2044 AA.  
AC P91620;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE STILL LIFE PROTEIN TYPE 2 (SIF TYPE 2).  
GN SIF.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Head;  
RX MEDLINE=97153054; PubMed=8999801;  
RA Sone M., Hoshino M., Suzuki E., Kuroda S., Kaibuchi K., Nakagoshi H.,  
Saigo K., Nabeshima Y.-I., Hama C.;  
RT "Still life, a protein in synaptic terminals of Drosophila homologous  
to GTP-GAP exchangers.";  
RL Science 275:543-547(1997).  
RN [2]  
RP ERRATUM.  
RA Sone M., Hoshino M., Suzuki E., Kuroda S., Kaibuchi K., Nakagoshi H.,  
Saigo K., Nabeshima Y.-I., Hama C.;  
RL Science 275:1405-1405(1997).  
CC -1- FUNCTION: REGULATES SYNAPTIC DIFFERENTIATION THROUGH THE  
ORGANIZATION OF ACTIN CYTOSKELETON POSSIBLY BY ACTIVATING RHO-LIKE  
GTPASES. IS LIKELY A FACTOR IN THE CASCADE OF RAC1 OR CDC42 IN THE  
NEURONS.  
CC -1- SUBCELLULAR LOCATION: LOCALIZES TO THE SUBMEMBRANOUS REGION OF  
SYNAPTIC TERMINALS.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; SIF TYPE 1 (AC P91621) AND SIF  
TYPE 2 (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- DEVELOPMENTAL STAGE: AT STAGE 14, EXPRESSION OCCURS IN EACH  
SEGMENT OF THE CENTRAL NERVOUS SYSTEM. AT STAGE 17, EXPRESSION  
BECOMES RESTRICTED TO THE SYNAPTIC REGIONS OF THE BRAIN AND  
VENTRAL NERVE CORD, WHERE SYNAPSES UNDERGO MATURATION.  
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.  
CC -1- SIMILARITY: CONTAINS 2 PH DOMAINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; D86546; BAA13108.1; -  
CC FlyBase; FBgn0019652; sif.  
CC InterPro; IPR001331; GDS\_CDC24.  
CC InterPro; IPR001478; PDZ.  
CC InterPro; IPR001849; PH.  
CC InterPro; IPR003116; RBD.

DR InterPro; IPR000219; RhoGEF.  
DR Pfam; PF00595; PDZ; 1.  
DR Pfam; PF00169; PH; 2.  
DR Pfam; PF02196; RBD; 1.  
DR Pfam; PF00621; RhoGEF; 1.  
DR SMART; SM00228; PDZ; 1.  
DR SMART; SM00233; PH; 2.  
DR SMART; SM00455; RBD; 1.  
DR SMART; SM00325; RhoGEF; 1.  
DR PROSITE; PS00741; GDS\_CDC24; 1.  
DR PROSITE; PS0106; PDZ; 1.  
DR PROSITE; PS0003; PH\_DOMAIN; 1.  
KW Guanine-nucleotide releasing factor; Developmental protein; Synapse;  
KW Repeat; Alternative splicing. 4 X 25 AA APPROXIMATE REPEAT.  
FT DOMAIN 62 249  
FT REPEAT 62 86  
FT REPEAT 94 118  
FT REPEAT 154 178  
FT REPEAT 225 249  
FT DOMAIN 819 937 PH 1.  
FT DOMAIN 1184 1273 PDZ.  
FT DOMAIN 1410 1673 DH.  
FT DOMAIN 1674 1767 PH 2.  
FT DOMAIN 467 470 POLY-PRO.  
FT DOMAIN 646 649 POLY-ARG.  
FT DOMAIN 1295 1298 POLY-PRO.  
FT DOMAIN 1898 1909 POLY-GLN.  
FT DOMAIN 1929 1933 POLY-PRO.  
SQ SEQUENCE 2044 AA; 228324 MW; 75D7CF21F49654B6 CRC64;  
  
Query Match 11.5%; Score 117.5; DB 1; Length 2044;  
Best Local Similarity 21.2%; Pred. No. 0.035;  
Matches 41; Conservative 42; Mismatches 99; Indels 11; Gaps 5;  
  
QY 1 IAKEVSTERTYUKDLEVIITSWFQSTVSKEDAMPEALKSLIFNPFELPKHFNFLKEIE 60  
Db 1412 VMVELDVTERTYVVKHLNLLLEHYLEPMKRETFLSNAEINALFGNIHEIVTFQGFQNLQNE 1471  
QY 61 QRLALMEGRSN-----AQIRDYQ-RIGDVMKNIQGMKHLAAHLWKHSEALELENGIKS 114  
Db 1472 ESDDL-EPDFNKFEGHCGFRNVLFAIGSAFLYYNHFKLYSSFCASHSKAKVLPN-EG 1529  
QY 115 SRLENFCRDFELQVKYCYPLNFTFLRPLRLHLMHYKQVLER---LCKHPPSHADFRCR 171  
Db 1530 NHALQEFLEARNPKQSHSSTLESYLIKPIQRIKYLKPLLLQOMRNLTDTRADEHVHLCAL 1589  
QY 172 AALAEITEMVAQL 184  
Db 1590 KGMKVAEHINEM 1602  
  
RESULT 11  
SIF1\_DROME STANDARD; PRT; 2064 AA.  
AC P91621;  
DT 15-JUL-1999 (Rel. 38, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE STILL LIFE PROTEIN TYPE 1 (SIF TYPE 1).  
GN SIF.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Head;  
RX MEDLINE=97153054; PubMed=8999801;  
RA Sone M., Hoshino M., Suzuki E., Kuroda S., Kaibuchi K., Nakagoshi H.,  
Saigo K., Nabeshima Y.-I., Hama C.;  
RT "Still life, a protein in synaptic terminals of Drosophila homologous





```
DR Pfam; PF00169; PH; 1.
DR SMART; SM00016; RhoGEF; 1.
DR SMART; SM00016; OPR; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00741; GDS_CDC24; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KW Guanine-nucleotide releasing factor; Nuclear protein.
FT DOMAIN 426 547 PH
FT DOMAIN 559 667 SER/THR-RICH.
SQ SEQUENCE 872 AA; 99101 MW; 726FAE33B519D69D CRC64;

Query Match 11.3%; Score 116; DB 1; Length 872;
Best Local Similarity 21.6%; Pred. No. 0.017;
Matches 46; Conservative 35; Mismatches 79; Indels 50; Gaps 6;

QY 1 IAEKSTERTYTKDLEVTISNQSVKSDAMPEALKSLIPNFPPLKHFHTNFKETLE 60
DB 232 VTAELVETELKYIQDLEVLISNYMVLQKQILSQDFTLS-IFTNLNILDQRRFLVGL 290
QY 61 QRLALWEGRSNAQIRYQRLGDVLMKNIQGMKHLAAHLWKHSEALENGIKSSRL 120
DB 291 MNLSL-----PVEQRGLGALFI-----ALEEGFSV---YQV 318
QY 121 FCRDF-----ELQKVCYL-----PLNTELLRPLRLMHYKQVLERLCKKHPPS 163
DB 319 FCTNFPNAQQLITDQNLKLVANLLEPSYELPALLIKPIQRIKCYKIPLLNQLLKTPSG 378
QY 164 HADFRDCRAALAEITEMVAQLHGTMIKMN 193
DB 379 YQVEELKQGMACVVRVANQVNETRIHEN 408

RESULT 13
ID GNRP_MOUSE STANDARD; PRT; 1262 AA.
AC P27671.
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE GUANINE NUCLEOTIDE RELEASING PROTEIN (GNRP) (RAS-SPECIFIC NUCLEOTIDE
DE EXCHANGE FACTOR CDC25) (CDC25MM).
GN RASGRF1 OR CDC25 OR GRF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C.
RX MEDLINE=93010996; PubMed=1396590;
RA Cen H., Lowy D.B.;
RT Isolation of multiple mouse cDNAs with coding homology to
RT Saccharomyces cerevisiae CDC25; Identification of a region related to
RT Bcr, Vav, Dbl and CDC24.;
RL EMBO J. 11:4007-4015(1992).
RN [2]
RP SEQUENCE OF 791-1262 FROM N.A.
RC STRAIN-SWISS; TISSUE=Brain;
RX MEDLINE=92289680; PubMed=1376246;
RA Martegani E., Vanoni M., Zippel R., Coccetti P., Brambilla R.,
RA Ferrarini C., Sturani E.P., Albergina L.;
RT Cloning by functional complementation of a mouse cDNA encoding a
RT homologue of CDC25, a Saccharomyces cerevisiae RAS activator.;
RL EMBO J. 11:2151-2157(1992).
RN [3]
RP SEQUENCE OF 1031-1226 FROM N.A.
RX MEDLINE=92357779; PubMed=1379731;
RA Wei W., Mosteller R.D., Saaval P., Gonzales E., McKinney D.,
RA Dasgupta C., Li P., Liu B.X., Broek D.;
RT Identification of a mammalian gene structurally and functionally
RT related to the CDC25 gene of Saccharomyces cerevisiae.;
```

```
RL Proc. Natl. Acad. Sci. U.S.A. 89:7100-7104(1992).
CC -!- FUNCTION: PROMOTES THE EXCHANGE OF RAS-BOUND GDP BY GTP.
CC -!- TISSUE SPECIFICITY: BRAIN.
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 RASGEF DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L20899; AAA02741.1; -.
DR EMBL; X59868; CAA42525.1; -.
DR PIR; S20730; S20730.
DR PIR; S22693; S22693.
DR MGD; MGI:99694; Rasgrf1.
DR InterPro; IPR001331; GDS_CDC24.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000651; RasGEFN.
DR InterPro; IPR001895; RasGRE_CDC25.
DR InterPro; IPR000219; RhoGEF.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00169; PH; 2.
DR Pfam; PF00617; RasGEF; 1.
DR Pfam; PF00618; RasGEFN; 1.
DR Pfam; PF00621; RhoGEF; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00147; RasGEF; 1.
DR SMART; SM00229; RasGEFN; 2.
DR SMART; SM00325; RhoGEF; 1.
DR PROSITE; PS00720; GDS_CDC25; 1.
DR PROSITE; PS00741; GDS_CDC24; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
DR PROSITE; PS50096; IQ; 1.
KW Guanine-nucleotide releasing factor.
FT DOMAIN 22 130
FT DOMAIN 208 233 IQ.
FT DOMAIN 248 459 DH.
FT DOMAIN 460 588 PH 2.
FT DOMAIN 1025 1259 RASGEF.
FT CONFLICT 1033 1033 E -> D (IN REF. 3).
SQ SEQUENCE 1262 AA; 144101 MW; 38BFE68F7C228DC8 CRC64;

Query Match 11.3%; Score 116; DB 1; Length 1262;
Best Local Similarity 21.6%; Pred. No. 0.027;
Matches 43; Conservative 34; Mismatches 90; Indels 32; Gaps 5;

QY 9 ERTYKDLKVITSWF-----QSTVSKEDAMPEALKSLIPNFPPLKHFHTNFKETQRL 63
DB 256 EAEYVOQLHLVNNFLRPLRMASSKKPPITHDDVSSIFLNSETINFLHOIFQGLKARI 315
QY 64 ALWEGRSNAQIRD-----YORIGDVLMKNIQGMKHLAAHLWKHSEALENGIKSS 115
DB 316 SSWPTLVLAADFLDILLPLMLNIYQE---FVRNHQYSLQILAH-----CKQN 357
QY 116 RLNEFCRDFELQKVC-YLPLNTFLRLPLRLMHYKQVLERLCKKHPPSHADFRCAAL 174
DB 358 RQDFKLLKQYKAPDCERTLETFTYPMFQIPRYILTLHELLAHPHEHVERNSLDYAK 417
QY 175 AEITEMVAQLHGTMIKMN 193
DB 418 SKLELSRIMHDEVSETN 436

RESULT 14
```

```
CC24_YEAST          STANDARD;          PRT;          854 AA.
ID  P11433;
AC  01-OCT-1989 (Rel. 12, Created)
DT  01-FEB-1996 (Rel. 33, Last sequence update)
DE  15-JUL-1999 (Rel. 38, Last annotation update)
DE  CELL DIVISION CONTROL PROTEIN 24 (CALCIUM REGULATORY PROTEIN).
GN  CDC24 OR CLS4 OR YAL041W.
OS  Saccharomyces cerevisiae (Baker's yeast).
OC  Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC  Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX  NCBI_TaxID=4932;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=87277425; PubMed=3301539;
RA  Miyamoto S., Ohya Y., Ohsumi Y., Anraku Y.;
RT  "Nucleotide sequence of the CLS4 (CDC24) gene of Saccharomyces
RL  cerevisiae.";
RL  Gene 54:125-132(1987).
RN  [2]
RP  REVISIONS, AND SIMILARITY TO CDC24 FAMILY.
RX  MEDLINE=92095962; PubMed=1755844;
RA  Miyamoto S., Ohya Y., Sano Y., Sakaguchi S., Iida H., Anraku Y.;
RT  "A DBL-homologous region of the yeast CLS4/CDC24 gene product is
RT  important for Ca(2+)-modulated bud assembly.";
RL  Biochem. Biophys. Res. Commun. 181:604-610(1991).
RN  [3]
RP  SEQUENCE FROM N.A.
RX  STRAIN=S288C / AB972;
RX  MEDLINE=95249563; PubMed=7731988;
RA  Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA  Hall J., Ouellette B.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA  Storms R.K.;
RT  "The nucleotide sequence of chromosome I from Saccharomyces
RT  cerevisiae.";
RL  Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC  -1- FUNCTION: PROMOTES THE EXCHANGE OF CDC42-BOUND GDP BY GTP.
CC  CONTROLS THE POLARITY OF CALMODULIN, AND THE CALCIUM REGULATORY
CC  PROCESS OF BUD EMERGENCE. CDC24 MAY BE INVOLVED IN THE INITIAL
CC  SELECTION AND ORGANIZATION OF THE BUDDING SITE.
CC  -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC  -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; M16809; AAA82871.1; -
DR  EMBL; U12980; AAC04990.1; -
DR  PIR; A27477; A27477.
DR  SGD; S0000039; CDC24.
DR  InterPro; IPR001715; Calponin_hom.
DR  InterPro; IPR001331; GDS_CDC24.
DR  InterPro; IPR000270; OPR.
DR  InterPro; IPR001849; PH.
DR  InterPro; IPR000219; RhoGEF.
DR  Pfam; PF00364; OPR; 1.
DR  Pfam; PF00621; RhoGEF; 1.
DR  SMART; SM00033; CH; 1.
DR  SMART; SM00016; OPR; 1.
DR  SMART; SM00233; PH; 1.
DR  SMART; SM00325; RhoGEF; 1.
DR  PROSITE; PS00741; GDS_CDC24; 1.
DR  PROSITE; PS50003; PH_DOMAIN; 1.
KW  Guanine-nucleotide releasing factor.
FT  DOMAIN 281 477 DH.
FT  DOMAIN 478 668 PH.
FT  DOMAIN 494 600 SER/THR-RICH.
FT  DOMAIN 681 778 SER/THR-RICH.
```

```
SQ  SEQUENCE 854 AA; 96939 MW; 856BCAF4EB0A67D1 CRC64;

Query Match 11.1%; Score 114; DB 1; Length 854;
Best Local Similarity 22.2%; Pred. No. 0.025;
Matches 49; Conservative 39; Mismatches 81; Indels 52; Gaps 9;

QY 1 IAKEVSTTERTYIKDLEVTISFQSTVSKEDAMPEALKSLIPNPFELKHFHTNFKLEIE 60
DB 282 IIEKVFATEYKVVHDLIEDKVRQQLDSNLTISEL-YMLFPNLGDAIDQRRFLISLE 340
QY 61 QRLALWEGRSNAQIR-DYQIRIGDVMLKNIQGMKHLAAHLWK-----HSEALEALENG 111
DB 341 -----INALVEPSKQIRGALFMH-----SKHFELKLYEPWSIGNAAIEFLSST 383
QY 112 IKSSR-RLENFCRDFELQVKCYLPNTLLRPLHRLMHYKQVLRCLKXHPHSHADFDR 169
DB 384 LHKMRVDESQRFIINKLE-----LQSFYKPKVQRLCRYPPLLVKELLAESSDDN-NTKE 436
QY 170 CRAAL-----AEITENVAQLHGTMTKMFNF 194
DB 437 LEAALDISKNIARSINENORRTENHQVVKLYGRVYVNWKG 477

RESULT 15
DBS_RAT ID DBS_RAT STANDARD; PRT; 937 AA.
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GUANINE NUCLEOTIDE EXCHANGE FACTOR DBS (DBL'S BIG SISTER) (MCF2
DE TRANSFORMING SEQUENCE-LIKE PROTEIN) (OST ONCOGENE) (FRAGMENT).
DE MCF2L OR OST.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Osteosarcoma;
RX MEDLINE=95045368; PubMed=7957046;
RA Horii Y., Beeler J.F., Sakaguchi K., Tachibana M., Miki T.;
RT "A novel oncogene, ost, encodes a guanine nucleotide exchange factor
RL that potentially links Rho and Rac signaling pathways.";
EMBO J. 13:4776-4786(1994).
CC -1- FUNCTION: GUANINE NUCLEOTIDE EXCHANGE FACTOR THAT POTENTIALLY
CC LINKS PATHWAYS THAT SIGNAL THROUGH RAC1, RHOA AND CDC42. CATALYZES
CC GUANINE NUCLEOTIDE EXCHANGE ON RHOA AND CDC42 AND INTERACTS
CC SPECIFICALLY WITH THE GTP-BOUND FORM OF RAC1, SUGGESTING THAT IT
CC FUNCTIONS AS AN EFFECTOR OF RAC1. MAY ALSO PARTICIPATE IN AXONAL
CC TRANSPORT IN THE BRAIN. BECOMES ACTIVATED AND HIGHLY TUMORIGENIC
CC BY TRUNCATION OF THE N-TERMINUS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN THE BRAIN, WHERE IT IS
CC FOUND IN NEURONS AND ALPHA-TANCYTES. LOWER LEVEL IN THE HEART.
CC -1- PTM: MAINLY PHOSPHORYLATED ON SERINE.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z35654; CAA84713.1; ALT_INIT.
DR InterPro; IPR001251; CRAL_TRIO.
DR InterPro; IPR001331; GDS_CDC24.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhoGEF.
```

```

DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00169; PH; 1.
DR SMART; SM00621; RhoGEF; 1.
DR SMART; SM00333; PH; 1.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00516; SEC14; 1.
DR SMART; SM00150; SPEC; 1.
DR PROSITE; PS00741; GDS.CDC24; 1.
DR PROSITE; PS50003; PH.DOMAIN; 1.
KW Guanine-nucleotide releasing factor; Proto-oncogene; Phosphorylation.
FT NON_TER 1
FT DOMAIN 499 502 POLY-GLU.
FT DOMAIN 514 692 DH.
FT DOMAIN 716 825 PH.
SQ SEQUENCE 937 AA; 106275 MW; 6A775209C305DFE5 CRC64;

Query Match 10.9%; Score 111; DB 1; Length 937;
Best Local Similarity 21.0%; Pred. No. 0.051;
Matches 43; Conservative 45; Mismatches 73; Indels 44; Gaps 9;

QY 1 IAKEVSTERTYKLDLEVTISWFQSTVSKEDAMPEAL-----KSLIFPNFELHKEHT 53
Db 511 VMNELDTERAYVEELLCVLEGYAAEMD-NPLMAHLISTGLQNKKNILFGNMEEIYHFHN 569
QY 54 N-FLKEIEQRLALWE--GRSNAQIRDYQRIQDVMLKNIQGMKHLAAHLWKHSEALEALEN 110
Db 570 RIFLELESCIDCELYGRCFELERMEEFQIYEKYCN-----KPRSESLWRQ----- 616
QY 111 GIKSSRLNFCRDFELQKVY-----LPLNTFLRLPLHLMHYKOVLERLCKKHPPSH 164
Db 617 -----CSDCPFFQECQKLDHKLSDSYLLKPVQRTIKYQLLLKEMILKY--SKH 663
QY 165 AD-FRDCRAALAEITENVAQLHGTM 188
Db 664 CEGAEDLQEALESILGLIKAVNDMSM 688

```

Search completed: December 6, 2001, 08:53:10  
Job time: 382 sec

GenCore version 4.5

Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:54:32 ; Search time 77.11 Seconds  
(without alignments)  
368.005 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_544\_737

Perfect score: 1023

Sequence: 1 IAKEVSTERYLKDLEVT.....AEITEMVAQLHGTMIKMFN 194

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL17.\*

1: sp-archaea.\*

2: sp-bacteria.\*

3: sp-fungi.\*

4: sp-human.\*

5: sp-invertebrate.\*

6: sp-mammal.\*

7: sp-mhc.\*

8: sp-organelle.\*

9: sp-phage.\*

10: sp-plant.\*

11: sp-rodent.\*

12: sp-virus.\*

13: sp-vertebrate.\*

14: sp-unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query	ID	Description
1	1023	100.0	1045	Q9Y4F1	Q9Y4F1 homo sapien
2	570	55.7	420	Q99K03	Q99K03 mus musculus
3	566.5	55.4	1054	Q94887	Q94887 homo sapien
4	320	31.3	562	Q9VN69	Q9VN69 drosophila
5	195.5	19.1	766	Q88387	Q88387 rattus norv
6	190.5	18.6	875	Q18372	Q18372 caenorhabdi
7	184.5	18.0	699	Q9P215	Q9P215 homo sapien
8	179	17.5	1676	Q9ULG4	Q9ULG4 homo sapien
9	179	17.5	1681	Q9NYG0	Q9NYG0 homo sapien
10	179	17.5	1696	Q9NZM3	Q9NZM3 homo sapien
11	176	17.2	411	Q9ER22	Q9ER22 rattus norv
12	176	17.2	493	Q18QX73	Q18QX73 rattus norv
13	176	17.2	516	Q43307	Q43307 homo sapien
14	172	16.8	1721	Q95216	Q95216 homo sapien
15	171	16.7	496	Q9NTG0	Q9NTG0 homo sapien
16	171	16.7	619	Q9HDC6	Q9HDC6 homo sapien
17	171	16.7	621	Q93504	Q93504 brachydanio
18	171	16.7	694	Q9UPP0	Q9UPP0 homo sapien
19	171	16.7	720	Q9NR80	Q9NR80 homo sapien

20	168.5	16.5	1658	11	Q9Z0R6	Q9Z0R6 mus musculus
21	166.5	16.3	1335	5	O00886	O00886 dictyosteli
22	166	16.2	820	5	O96035	O96035 ciona savig
23	166	16.2	1721	4	Q9UNK2	Q9UNK2 homo sapien
24	163	15.9	1714	11	Q9Z0R4	Q9Z0R4 mus musculus
25	157.5	15.4	1737	5	Q9TW28	Q9TW28 dictyosteli
26	157	15.3	698	4	Q9NXY1	Q9NXY1 homo sapien
27	155	15.2	727	11	O88841	O88841 mus musculus
28	154	15.1	647	5	O9VS45	O9VS45 drosophila
29	152	14.9	733	11	O88842	O88842 mus musculus
30	149	14.6	1522	4	O15085	O15085 homo sapien
31	147	14.4	1527	11	O9ES67	O9ES67 rattus norv
32	138	13.5	761	3	Q9HF59	Q9HF59 ashbya goss
33	131	12.8	1544	4	Q9NZN5	Q9NZN5 homo sapien
34	129.5	12.7	985	11	Q9ESG7	Q9ESG7 mus musculus
35	129	12.6	618	11	Q9CWR0	Q9CWR0 mus musculus
36	128.5	12.6	1028	4	Q9UGQ4	Q9UGQ4 homo sapien
37	128	12.5	1101	5	Q9VDF9	Q9VDF9 drosophila
38	127.5	12.5	1011	5	Q9W0R9	Q9W0R9 drosophila
39	127.5	12.5	1011	5	Q9GTU9	Q9GTU9 drosophila
40	126.5	12.4	750	4	O15086	O15086 homo sapien
41	125.5	12.3	958	4	Q9H023	Q9H023 homo sapien
42	125	12.2	594	5	Q9VS95	Q9VS95 drosophila
43	124.5	12.2	793	5	Q9NHV9	Q9NHV9 drosophila
44	123.5	12.1	852	5	Q9VWJ5	Q9VWJ5 drosophila
45	123	12.0	630	11	Q9ES27	Q9ES27 mus musculus

## ALIGNMENTS

RESULT 1

ID	Q9Y4F1	PRELIMINARY;	PRT;	1045 AA.
AC	Q9Y4F1;			
DT	01-NOV-1999 (TReMBLrel. 12, Created)			
DT	01-NOV-1999 (TReMBLrel. 12, Last sequence update)			
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)			
DE	CDEP.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=CARTILAGE;			
RX	MEDLINE=98086358; PubMed=9425278;			
RA	Koyano Y., Kawamoto T., Shen M., Yan W., Noshiro M., Fujii K.,			
RA	Kato Y.;			
RT	*Molecular cloning and characterization of CDEP, a novel human protein			
RT	containing the ezrin-like domain of the band 4.1 superfamily and the			
RL	Dbl homology domain of Rho guanine nucleotide exchange factors.;			
RL	Biochem. Biophys. Res. Commun. 241:369-375(1997).			
DR	EMBL; AB008430; BAA24267.1; .			
DR	HSSP; P08567; iPLS.			
DR	InterPro; IPR000299; Band_4.1.			
DR	InterPro; IPR001849; PH.			
DR	InterPro; IPR000219; RhoGEF.			
DR	Pfam; PF00373; Band_41; 1.			
DR	Pfam; PF00169; PH; 2.			
DR	Pfam; PF00621; RhoGEF; 1.			
DR	PRINTS; PR00935; BAND41.			
DR	SMART; SM00295; B41; 1.			
DR	SMART; SM00233; PH; 2.			
DR	SMART; SM00325; RhoGEF; 1.			
DR	PROSITE; PS00660; BAND_41_1; UNKNOWN_1.			
DR	PROSITE; PS00037; BAND_41_3; 1.			
DR	PROSITE; PS50003; PH_DOMAIN; 2.			
SQ	SEQUENCE 1045 AA; 118632 MW; 0E8B2D61C0F58417 CRC64;			

Query Match 100.0%; Score 1023; DB 4; Length 1045;  
Best Local Similarity 100.0%; Pred. No. 1.4e-81;

```
Matches 194; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 IAKVSTTERTYKDLVITSWFQSTVSKEDAMPEALKSLIFPNFPLHFKHFTNFKLEIE 60
RL IAKVSTTERTYKDLVITSWFQSTVSKEDAMPEALKSLIFPNFPLHFKHFTNFKLEIE 60
DB 544 IAKVSTTERTYKDLVITSWFQSTVSKEDAMPEALKSLIFPNFPLHFKHFTNFKLEIE 603
QY 61 ORLALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLEN 120
RL ORLALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLEN 120
DB 604 ORLALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLEN 663
QY 121 FCRDFELQKVCYLPNTFLLRPLHRLMHYKQVLERLCKHPPSHADFRDCRAALAEITEM 180
RL FCRDFELQKVCYLPNTFLLRPLHRLMHYKQVLERLCKHPPSHADFRDCRAALAEITEM 180
DB 664 FCRDFELQKVCYLPNTFLLRPLHRLMHYKQVLERLCKHPPSHADFRDCRAALAEITEM 723
QY 181 VAQLHGTMIKMENF 194
RL VAQLHGTMIKMENF 194
DB 724 VAQLHGTMIKMENF 737

RESULT 2
Q99KU3 ID Q99KU3 PRELIMINARY; PRT; 420 AA.
AC Q99KU3;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE UNKNOWN (PROTEIN FOR IMAGE:3493093) (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC004009; AAH04009.1; -.
FT NON_TER 1
SQ SEQUENCE 420 AA; 48387 MW; 804411A6E6613F96 CRC64;

Query Match 55.7%; Score 570; DB 11; Length 420;
Best Local Similarity 94.8%; Pred. No. 3e-42;
Matches 106; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
QY 83 VMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLENFCRDFELQKVCYLPNTFLLRP 142
DB 1 VMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLENFCRDFELQKVCYLPNTFLLRP 60
QY 143 LHRMLHYKQVLERLCKHPPSHADFRDCRAALAEITEMVAQLHGTMIKMENF 194
DB 61 LHRMLHYKQVLERLCKHPPSHADFRDCRAALAEITEMVAQLHGTMIKMENF 112

RESULT 3
Q94887 ID Q94887 PRELIMINARY; PRT; 1054 AA.
AC Q94887;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE KIAA0793 PROTEIN.
GN KIAA0793.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RA SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
```

```
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.
DNA Res. 5:277-286(1998).
[2]
RP SEQUENCE OF 1-754 FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence."
Genome Res. 8:1097-1108(1998).
[3]
RP SEQUENCE OF 1-754 FROM N.A.
RA Sun H., Stoneking T., Langston Y., LaPlant Y.;
RT "The sequence of Homo sapiens BAC clone RG442F18."
Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
[4]
RP SEQUENCE OF 1-754 FROM N.A.
RA Waterston R.H.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.
[5]
RP SEQUENCE OF 1-754 FROM N.A.
RA Waterston R.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
[6]
RP SEQUENCE OF 1-754 FROM N.A.
RA Waterston R.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AB018336; BAA34513.1; -.
DR EMBL; AC005104; AAD12224.1; -.
DR HSSP; P08567; IPLS.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR001849; PH.
DR InterPro; IPR002119; RhogEF.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00169; PH; 2.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR SMART; SM00233; PH; 2.
DR SMART; SM00325; RhogEF; 1.
DR PROSITE; PS00660; BAND_41_1; UNKNOWN_1.
DR PROSITE; PS50057; BAND_41_3; 1.
DR PROSITE; PS50003; PH_DOMAIN; 2.
SQ SEQUENCE 1054 AA; 119887 MW; 7DC279F69A307E3A CRC64;

Query Match 55.4%; Score 566.5; DB 4; Length 1054;
Best Local Similarity 53.1%; Pred. No. 1.7e-41;
Matches 103; Conservative 36; Mismatches 54; Indels 1; Gaps 1;
QY 1 IAKVSTTERTYKDLVITSWFQSTVSKEDAMPEALKSLIFPNFPLHFKHFTNFKLEIE 60
DB 539 IAKVSTTERTYKDLVITSWFQSTVSKEDAMPEALKSLIFPNFPLHFKHFTNFKLEIE 598
QY 61 ORLALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLE 119
DB 599 ORLALWEGRSNAQIRDYORIGDVMLKNIQGMKHLAAHLWKHSEALENGIKSSRRLE 658
QY 120 NFCRDFELQKVCYLPNTFLLRPLHRLMHYKQVLERLCKHPPSHADFRDCRAALAEITE 179
DB 659 AVYKEFELOKVCYLPNTFLLRPLHRLMHYKQVLERLCKHPPSHADFRDCRAALAEITE 718
QY 180 MVAQLHGTMIKMEN 193
DB 719 VTTTLQHILIRLEN 732

RESULT 4
Q9VNV6 ID Q9VNV6 PRELIMINARY; PRT; 562 AA.
AC Q9VNV6;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
```

```

QY      121 FCRDFELQKVCYLPLNTFTLRPLRLHLMHYKVQLERLCKHHPPSHADFRCRA 172
       :::: |||||: ||:||:||||: |||||: | |: |
Db      209 VYKEFEQKVCYLPIGELLKPLNRLLYLIERLCDDVYGEHIDYADAMA 260

RESULT      5
O88387 PRELIMINARY; PRT; 766 AA.
ID AC O88387;
DC DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE ACTIN-FILAMENT BINDING PROTEIN FRABIN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=101116;
RN [1]
RS SEQUENCE FROM N.A.
RX MEDLINE=98334590; PubMed=9668039;
RA Obaishi H., Nakanishi H., Mandai K., Satoh K., Satoh A., Takahashi K.
RA Miyahara M., Nishiooka H., Takaishi K., Takai Y.:
RT "Frabin, a novel FGDI-related actin filament-binding protein capable
RT of changing cell shape and activating c-Jun N-terminal kinase.";
RL J. Biol. Chem. 273:18697-18700(1998).
DR EMBL: AF038388; AAC27698.1; -.
DR InterPro: IPR000219; RhogEF.
DR InterPro: IPR000306; Znf_FYVE.
DR InterPro: IPR000822; Znf-C2H2.
DR InterPro: IPR001849; PH.
DR Pfam: PF00169; PH; 2.
DR Pfam: PF00621; RhogEF; 1.
DR Pfam: PF01363; FYVE; 1.
DR PROSITE: PS50003; PH_DOMAIN; 2.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_N_1.
DR SMART: SM00064; FYVE; 1.
DR SMART: SM00233; PH; 2.
DR SMART: SM00325; RhogEF; 1.
SQ SEQUENCE 766 AA; 86453 MW; E34831DC0B0B334C CRC64;

Query Match 19.1%; Score 195.5; DB 11; Length 766;
Best Local Similarity - 26.3%; Pred. No. 4.3e-09;
Matches 52; Conservative 43; Mismatches 90; Indels 13; Gaps

QY      1 IAEVSTERTYTKDLEVTISWFSQVSKE--DAMPEALKSLIFPNFELHKFHTNF-L 56
       ||: |||: |::|: |::|: |::|: |::|: |::|: |::|: |
Db      210 IATELLTIRAYSRNLNDQVFYKLLLEANRGSPFAEMVNKIFSINSINAFHSKFLL 269

QY      57 KEIQRALMEGRSNAQIRDYQRIGDVMLKNIOGMKHLAAHMKHSEALEALENGIKSSR 116
       ||: ||: || |||||: |::|: |::|: |::|: |::|: |
Db      270 PELEKRMQEW-----TTPRIGDILQKLAPFLNKGYGVKGFDNAVLYKNMTERVP 321

QY      117 RLENFCRDFELQKVC-YLPLNTFTLRPLRLHLMHYKVQLERLCKHHPPSHADFRCRAA 175
       ::: ||: |::|: |::|: |::|: |::|: |::|: |::|: |
Db      322 QFKSVTEIEIQKICGSLTLQHMHMLEPIQRIPYEMLLDKYLKLLSPDAPDNDAKKSL 381

QY      176 EITEMVAQLHGTMKWEN 193
       |: : : : ||||
Db      382 IIISTAASHNSAIRKMEN 399

RESULT      6
O18372 PRELIMINARY; PRT; 875 AA.
ID AC O18372;
DC DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE C33D9.1 PROTEIN.
DS C33D9.1.
OS Caenorhabditis elegans.
```

```

Query Match      31.3%; Score 320; DB 5; Length 562;
Best Local Similarity 38.4%; Pred. No. 3.6e-20;
Matches 66; Conservative 35; Mismatches 65; Indels 6; Gaps 3;

QY      1 IAEVSTERTYKDLVITSWFSTVSKEDAMPEALKSLIFPNPEPLHKHTNPLKIE 60
      :|||: ||||| |||:|: |: | | | | | | | | | | | | | | | |
Db      95 LAKELLMTERTYKDLVLTNTRFQVLSLGDV--EQQLPL-FELLDSLAQHNLFLRDIE 151

QY      61 QRLALWGRSNAQRDYQRTGDVYMKNIQGMKHHAAHLWKHSEALEALENGICKSRRLLEN 120
      |: ||||| : ||||| |||:|: |: | | | | | | | | | | | | | | | |
Db      152 HRMVOEWGRG---HEAHRIGDVMNKAHMAPIYDEVYVOTHLDIHGMNDWMYEGDERP 208

```





```

DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00168; C2; 1.
DR Pfam: PF00036; ehfand; 2.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhoGEF; 1.
DR Pfam: PF00018; SH3; 5.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhoGEF; 1.
DR Pfam: PF00018; SH3; 5.
DR PRINTS: PR00499; P67PHOX.
DR PRINTS: PR00452; SH3DOMAIN.
DR SMART; SM00239; C2; 1.
DR SMART; SM00054; Efh; 2.
DR SMART; SM00027; EH; 2.
DR SMART; SM00233; PH; 1.
DR SMART; SM00027; EH; 2.
DR SMART; SM00325; RhoGEF; 1.
DR SMART; SM00326; SH3; 5.
DR PROSITE; PS00004; C2_DOMAIN_2; 1.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE; PS00003; PH_DOMAIN; 1.
DR PROSITE; PS00002; SH3; 5.
KW Calcium-binding.
SQ SEQUENCE 1676 AA; 191178 MW; B877291582FB7F99 CRC64;

Query Match 17.5%; Score 179; DB 4; Length 1676;
Best Local Similarity 25.8%; Pred. No. 2.9e-07;
Matches 51; Conservative 47; Mismatches 80; Indels 20; Gaps 5;

QY 4 EVSTTERTYKLDLVITSFQSTVSKEDAMPEALKSLIFPNPEPLHKFHTNFKLEIQRL 63
DB 1195 ELTIQTEERYMADQLVVEVFKRMAESGFLTEGEMALIFVNMKELIMSNKLLKALVR- 1253
QY 64 ALWEGSRNAQIRDYQRIGDVMLKNIQGMKHLAAHLWKHSEALEENGKSSRRLENFCR 123
DB 1254 ----KTGGEKMPVMIGDILAAE---LSHMQAYIRFCSCQL-----NGAALQOKTDEDT 1302
QY 124 DEF--LQKVCY-----LPLNTFLLRPLRLMHYKOVLERLCKKHPPSHADFRDCRAALA 175
DB 1303 DKFEFLKKLASDPRCKGMPLSSFLPKPMQIRITRYPLLRISILENTPESHADHSSKLAL 1362
QY 176 EITEMVAQLHGTMKMN 193
DB 1363 RAEELCSQVNEGVRKEN 1380

RESULT 9
Q9NYG0 PRELIMINARY; PRT; 1681 AA.
AC Q9NYG0;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE INTERSECTIN 2.
GN SH3DLB.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Seifert M., Engel M., Welter C.;
RT "Intersectin 2 (SH3DLB), human homolog of mouse Ecs2 protein.";
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO EF-HAND FAMILY.
DR EMBL; AF248540; AAF63600.1; -.
DR InterPro: IPR000008; C2.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000261; EPS15_repeat.
DR InterPro: IPR000108; Neu_cyt_fact_2.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR001452; SH3.

QY 9 EYSTTERTYKLDLVITSFQSTVSKEDAMPEALKSLIFPNPEPLHKFHTNFKLEIQRL 63
DB 1195 ELTIQTEERYMADQLVVEVFKRMAESGFLTEGEMALIFVNMKELIMSNKLLKALVR- 1253
QY 64 ALWEGSRNAQIRDYQRIGDVMLKNIQGMKHLAAHLWKHSEALEENGKSSRRLENFCR 123
DB 1254 ----KTGGEKMPVMIGDILAAE---LSHMQAYIRFCSCQL-----NGAALQOKTDEDT 1302
QY 124 DEF--LQKVCY-----LPLNTFLLRPLRLMHYKOVLERLCKKHPPSHADFRDCRAALA 175
DB 1303 DKFEFLKKLASDPRCKGMPLSSFLPKPMQIRITRYPLLRISILENTPESHADHSSKLAL 1362
QY 176 EITEMVAQLHGTMKMN 193
DB 1363 RAEELCSQVNEGVRKEN 1380

RESULT 10
Q9NZM3 PRELIMINARY; PRT; 1696 AA.
AC Q9NZM3;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE INTERSECTIN 2 LONG ISOFORM.
GN ITS2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Pucharcos C., Estivill E., de la Luna S.;
RT "Cloning and characterization of human intersectin 2: a protein
involved in endocytosis.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO EF-HAND FAMILY.
DR EMBL; AF182198; AAF59903.1; -.
DR InterPro: IPR000008; C2.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000261; EPS15_repeat.
DR InterPro: IPR000108; Neu_cyt_fact_2.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhoGEF.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00168; C2; 1.
DR Pfam: PF00036; ehfand; 2.
DR Pfam: PF00169; PH; 1.

```

[illegible]

	RC	TISSUE=BRAIN;
	RX	MEDLINE=99017974; PubMed=9799604;
	RA	Guipponi M., Scott H.S., Chen H., Schebesta A., Rossier C.,
	RR	Antonarakis S.E.;
	RT	"Two isoforms of a human intersectin (ITSN) protein are produced by
	RT	brain-specific alternative splicing in a stop codon.";
	RL	Genomics 53:369-376(1998).
	CC	-I- SIMILARITY: TO EF-HAND FAMILY.
	DR	EMLB; AF064244; AAC78611.1; -.
	DR	HSSP; P29355; ISEM.
	DR	InterPro: IPRO00008; C2.
	DR	InterPro: IPRO02048; EF-hand.
	DR	InterPro: IPRO0261; EFS15_repeat.
	DR	InterPro: IPRO00108; Neu_cyt_fact_2.
	DR	InterPro: IPRO01849; PH.
	DR	InterPro: IPRO00219; RhogEF.
	DR	InterPro: IPRO01452; SH3.
	DR	Pfam; PF00168; C2; 1.
	DR	Pfam; PF00036; ehand; 3.
	DR	Pfam; PF00169; PH; 1.
	DR	Pfam; PF00621; RhogEF; 1.
	DR	Pfam; PF00018; SH3; 5.
	DR	PRINTS; PR00499; P67PHOX.
	DR	PRINTS; PR00452; SH3DOMAIN.
	DR	SMART; SM00239; C2; 1.
	DR	SMART; SM00054; Eph; 2.
	DR	SMART; SM00027; EH; 2.
	DR	SMART; SM00233; PH; 1.
	DR	SMART; SM00325; RhogEF; 1.
	DR	SMART; SM00326; SH3; 5.
	DR	PROSITE; PS50004; C2_DOMAIN_2; 1.
	DR	PROSITE; PS00018; EF_HAND; UNKNOWN_2.
	DR	PROSITE; PS50003; PH_DOMAIN; 1.
	DR	PROSITE; PS50002; SH3; 5.
KW		Calcium-binding.
SQ		SEQUENCE 1721 AA; 195559 MW; 851A5CFB2BC4EBFC CRC64;
		Query Match 16.8%; Score 172; DB 4; Length 1721;
		Best Local similarity 20.9%; Pred.No.1.2e-06;
		Matches 40; Conservative 52; Mismatches 93; Indels 6; Gaps
QY	4	EVSTTERTYLKDELVITVSFQTGSKEDAMPALKSLIPFPPEPLHKHFNFLKETEQRL 63   :   :   ::     : : :   :      :   :   :   :   :   :
Dd	1244	ELIVTEENVNDLQLVTETFFQPKLMSELLEKEVAMIFVNKKELINCMNIKLLALRV- 1302
QY	64	ALWEGRSNAQIRDYQRIGDVMLNKGKHAAHLWKHSEALENGIKSSRLENFCR 123   :   :   ::     :   :   :   :   :   :   :   :   :
Dd	1303	----KMSGEKPVKMGIIDLSAQLPHPQIRFCSROLNGAALIQQTKDEAPDFEVR 1358
QY	124	DFELOKVCY-LPINTFLRLPRLHMYKOVLRIERLKCHPPSHADFRCAALALETWVA 182   :   :   ::     :   :   :   :   :   :   :   :   :
Dd	1359	RLEMDPRCKMPGLSSLFKPMQGVRYTPLIKNILEENTPENHPDHSHLHALEKAELCS 1418   :   :   ::     :   :   :   :   :   :   :   :   :
QY	183	QLHGTMKMEN 193   : : :   :
Dd	1419	QVNEGVREKEN 1429
RESULT 15		
Q9NTG0		
ID	Q9NTG0	PRELIMINARY; PRT; 496 AA.
AC	CANTGO;	
DT	01-OCT-2000 (TREMBLrel. 15, Created)	
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)	
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)	
GN	HYPOPHYSICAL 57.7 KDA PROTEIN (FRAGMENT).	
DE	DKFZP343G2016.	
OS	Homo sapiens (Human).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
ON	NCBI_Taxid=9606;	
RX	[1]	



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:51:40 ; Search time 72.4 Seconds  
(without alignments)  
382.644 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_1\_374

Perfect score: 1956

Sequence: 1 MGEIEQRPTGSLGAPENS.....GHKKVOPERKSHIRSL 374

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*

- 1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*
- 2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*
- 3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*
- 4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*
- 5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*
- 6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*
- 7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*
- 8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*
- 9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*
- 10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*
- 11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*
- 12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*
- 13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*
- 14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*
- 15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*
- 16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*
- 17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*
- 18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*
- 19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*
- 20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*
- 21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*
- 22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1956	100.0	1045	20	Human chondrocyte-
2	1952	99.8	1045	21	Human cytoskeleton
3	689.5	35.3	1005	21	Human cytoskeleton
4	587.5	30.0	687	22	Human protein sequ
5	506.5	25.9	913	18	Protein tyrosine p
6	506.5	25.9	913	20	Human TPPI1 protei
7	399.5	20.4	369	22	Human protein tyro
8	359	18.4	635	21	Human colon cancer
9	356	18.2	1174	19	Human protein tyro
10	356	18.2	1174	20	Human protein tyro
11	348	17.8	586	20	Amino acid sequenc

12	328	16.8	590	20	AAW94458	Human neurofibroma
13	328	16.8	591	18	AAW09648	Human merlin prote
14	328	16.8	595	15	AAW60398	Merlin protein enc
15	328	16.8	596	20	AAW94459	Human neurofibroma
16	327	16.7	584	20	AAW94457	Mouse neurofibroma
17	327	16.7	591	20	AAW94456	Mouse neurofibroma
18	327	16.7	596	20	AAW94455	Mouse neurofibroma
19	326.5	16.7	305	22	AAW25892	Human protein sequ
20	325	16.6	584	18	AAW09647	Mouse merlin prote
21	323	16.5	591	18	AAW09646	Mouse merlin prote
22	323	16.5	596	18	AAW09645	Mouse merlin prote
23	320.5	16.4	495	22	AAW41770	Human polypeptide
24	312	16.0	445	22	AAW39984	Human polypeptide
25	286.5	14.6	279	22	AAW25733	Human protein sequ
26	266	13.6	450	22	AAW2607	Human protein sequ
27	265	13.5	2466	16	AAW71498	Human protein tyro
28	265	13.5	2466	19	AAW75999	Intracellular prot
29	265	13.5	2466	21	AAW90272	Human TPPI1 phosph
30	265	13.5	2485	21	AAW19343	Amino acid sequenc
31	245	12.5	142	22	AAW74414	Human colon cancer
32	223.5	11.4	1105	22	AAW25567	Human protein sequ
33	166	8.5	1007	22	AAW50660	C. elegans UNC-5 p
34	164	8.4	184	22	AAW25248	Human protein sequ
35	157	8.0	436	22	AAW73954	Human colon cancer
36	141	7.2	91	22	AAW13919	Peptide #353 encod
37	141	7.2	91	22	AAW26325	Peptide #362 encod
38	141	7.2	91	22	AAW01662	Peptide #344 encod
39	128	6.5	92	21	AAW01312	Human secreted pro
40	124	6.3	94	21	AAW01151	Human secreted pro
41	123.5	6.3	348	21	AAW42166	Human ORFX ORF1930
42	121.5	6.2	91	21	AAW00826	Human secreted pro
43	119	6.1	97	21	AAW56466	Human prostate can
44	115	5.9	2541	21	AAW41087	Human ORFX ORF851
45	112	5.7	69	22	AAW36561	Peptide #10598 enc

ALIGNMENTS

RESULT 1  
AAW07482  
ID AAY07482 standard; Protein; 1045 AA.  
XX  
AC AAY07482;  
XX  
DT 17-AUG-1999 (first entry)  
XX  
DE Human chondrocyte-derived protein CDEP.

XX  
KW Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;  
KW Db1 homology domain; pleckstrin homology domain; rheumatoid 'arthritis;  
KW drug.  
XX  
OS Homo sapiens.  
XX  
PN WO928458-A1.  
XX  
PD 10-JUN-1999.  
XX  
PF 27-NOV-1998; 98WO-JP05348.  
XX  
PR 27-NOV-1997; 97JP-0342060.

XX  
PA (CHUS ) CHUGAI SEIYAKU KK.  
XX  
PI Kato Y, Kawamoto T, Koyano Y;  
XX  
DR WPI; 1999-371117/31.  
XX  
DR N-PSDB; AAW79183.  
XX  
PT Protein CDEP expressed in differentiated chondrocytes, and gene  
XX encoding it

```
PS Claim 2; Fig 1; 59pp; Japanese.
XX
CC This sequence represents a protein (CDEP) expressed in differentiated
CC human foetal chondrocytes, which contains an ezrin-like domain, a Dbl
CC homology (DH) domain and a pleckstrin homology (PH) domain. The encoding
CC nucleic acid or protein can be used in the investigation and treatment of
CC cancers and arthritic diseases (including chronic rheumatoid arthritis),
CC or for screening of candidate anticancer drugs.
XX
SQ Sequence 1045 AA;

Query Match      100.0%; Score 1956; DB 20; Length 1045;
Best Local Similarity 100.0%; Pred. No. 1.1e-190;
Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGIEQRPPTGSRIGAPENSGISTLERGQKPPPTPSGKLVSIQIMLDDTQEAFFVQRA 60
Db 1 mgeiedrptpsrlgapensgistlergqkppptpsgklvsiqlmlddtqea fvpqra 60
QY 61 PGKVLIDAVCNHLNVEGDYFGLFEPDHHKITVWLDLLKPIVKQIRRPKHVVVVKFF 120
Db 61 pgkvlidavcnhlnvegd yfglef dphkkitvwldllkpi vkqirrp khvvvkvkff 120
QY 121 PPDHTQLQEELTRYLFALQVKQDLAQGRLCNDSAAALLSHIVQSEIGDFDEALDREHL 180
Db 121 ppdhtqlqeeltrylfalqvkqdlagrltcndtsaallshivqseigdfdealdrehl 180
QY 181 AKNKYIPQDALEDKIVEFHNNHIGQTPAESDFQLLEIARRLEMYGIRLHPAKDRGCTKI 240
Db 181 aknkyipqdaledkivefhnnhigdt paesdfqlleiar rlemygir lhpakdregtki 240
QY 241 NLAVANTGILVFGQFTKINAFNNAKVRKLSFKRKRFLIKLRPDANSAYQDTLEFLMASRD 300
Db 241 nlavantgilvfgftkinafnnavkrklsfkrkrfliklrpdansayqdtleflmasrd 300
QY 301 FCKSFWKICVEHAFRLFEPPKPKPVLFSGSSFRSGRTQKQVLDVYKGGHKKVQ 360
Db 301 fcksfwkicvehafrrlfeepkpkpvlfsgssfrsgrtqkqvldvykegg hkkvq 360
QY 361 FERKSHKHSIRSL 374
Db 361 ferkshkhsirsl 374

RESULT 2
AAY91947
ID AAY91947 standard; Protein; 1045 AA.
XX
AC AAY91947;
XX
DT 19-JUL-2000 (first entry)
XX
DE Human cytoskeleton associated protein 2 (CYSKP-2).
XX
KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;
KW autoimmunity; inflammatory, vesicle trafficking; neurological;
KW cardiovascular; cell motility; reproductive; muscle disorder.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Modified-site 23
FT /note= "potential phosphorylation site"
FT Modified-site 36
FT /note= "potential phosphorylation site"
FT Modified-site 41
FT /note= "potential phosphorylation site"
FT Domain 47..85
FT /note= "signature sequence"
FT Modified-site 92
FT /note= "potential phosphorylation site"
FT Domain 94..123
FT /note= "potential phosphorylation site"

/note= "signature sequence"
144..190
/note= "signature sequence"
150
/note= "potential phosphorylation site"
152
/note= "potential N-glycosylation site"
196..249
/note= "signature sequence"
207
/note= "potential phosphorylation site"
261..279
/note= "signature sequence"
270
/note= "potential phosphorylation site"
336
/note= "potential phosphorylation site"
340
/note= "potential phosphorylation site"
343
/note= "potential phosphorylation site"
366
/note= "potential phosphorylation site"
370
/note= "potential phosphorylation site"
396
/note= "potential phosphorylation site"
408
/note= "potential phosphorylation site"
418
/note= "potential phosphorylation site"
448
/note= "potential phosphorylation site"
495
/note= "potential N-glycosylation site"
525
/note= "potential phosphorylation site"
538
/note= "potential phosphorylation site"
542
/note= "potential phosphorylation site"
549..
/note= "potential phosphorylation site"
551
/note= "potential phosphorylation site"
571
/note= "potential phosphorylation site"
657
/note= "potential phosphorylation site"
658
/note= "potential phosphorylation site"
706
/note= "potential phosphorylation site"
770
/note= "potential phosphorylation site"
789
/note= "potential phosphorylation site"
811
/note= "potential phosphorylation site"
815
/note= "potential phosphorylation site"
826
/note= "potential phosphorylation site"
839
/note= "potential phosphorylation site"
770..773
/note= "signature sequence"
840
/note= "potential phosphorylation site"
842
/note= "potential phosphorylation site"
872
/note= "potential phosphorylation site"
/note= "potential phosphorylation site"
```

FT	Modified-site	878	/note=	"potential phosphorylation site"	
FT	Modified-site	883	/note=	"potential phosphorylation site"	
FT	Modified-site	889	/note=	"potential phosphorylation site"	
FT	Modified-site	898	/note=	"potential phosphorylation site"	
FT	Modified-site	919	/note=	"potential N-glycosylation site"	
FT	Modified-site	923	/note=	"potential phosphorylation site"	
FT	Modified-site	966	/note=	"potential phosphorylation site"	
XX	WO200017355-A2.				
PN	30-MAR-2000.				
XX	17-SEP-1999;	99WO-US21565.			
XX	18-SEP-1998;	98US-0172226.			
PR	27-APR-1999;	99US-0131321.			
XX	(INCY-) INCYTE PHARM INC.				
PA	Lal P. Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;				
PI	Guegler KJ, Patterson C, Azimzai Y, Baughn MR;				
XX	WPI; 2000-283582/24.				
DR	N-PSDB; AAA08582.				
XX	Human cytoskeleton associated proteins, used to treat cell				
PT	proliferative, autoimmune/inflammatory, vesicle trafficking,				
PT	neurological, cell motility, reproductive and muscle disorders				
XX	Claim 1; Page 82-84; 113pp; English.				
XX	AAV91946-61 show human cytoskeleton associated proteins 1 to 1				
CC	to CYSKP-16) respectively. The sequences can be used to treat				
CC	diagnose cancer and cell proliferative, autoimmune/inflammatory				
CC	trafficking, neurological, cardiovascular, cell motility, repr				
CC	and muscle disorders. Pharmaceutical compositions containing				
CC	CYSKP-16 can be used to treat or prevent disorders associated				
CC	decreased expression or activity of CYSKP (claimed), for exam				
CC	atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis				
CC	cancers, autoimmune/antiinflammatory disorders such as allergi				
CC	asthma, acquired immunodeficiency syndrome (AIDS), Crohn's di				
CC	diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scl				
CC	and trauma. CYSKP antagonists can be used to treat or prevent				
CC	disorder associated with increased expression or activity of C				
CC	(claimed).				
XX	Sequence 1045 AA;				
SQ					

Query Match	99.8%	Score 1953;	DB 21;	Length 1045;
Best Local Similarity	99.7%	Pred. No. 2.7e-190;		
Matches 373; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

Qy	1	MGEIEQRPTPGSRIGAPENSIGISLERGQKPPPTPSGKLVSIKIQMLDDTQEAFAFVQRA	60
Db	1	mgeieqrtpgsrli gapensigistlergkppptpsgklvsiqimlddtqea fvpqra	60
Qy	61	PGKVLLDAVCNHLNLVEGDYGLGFEPDHHKTTVWLDDLKPIVKQTRRPKHVVVVKFWVXFF	120
Db	61	pgkvlldavcnhlnlvegdylgfglefphdhkktvwldllkpkvgirrpkhvvvkvfwxkf	120
Qy	121	PPDHTQLEETRYLFAIQVKQDLAQGRLLTCDNTSAAALLSHIVQSETGDPDEALDREHL	180
Db	121	ppdhtqleetrylfa lqvkdlaqgrlltcdntsaallshivqse tgdpedeal drehl	180
Qy	181	AKNKYIPQQDALEPKIVFEFHINHGIGQTPAESDFQLLEIARLEMYGIRLHPAKDREGTKI	240

Db	181	aknyvipqgdaledkivelfhnhhigqtpaesdfqlleiarrienygirihpakdregtk1	240
Qy	241	nlavantgilvfogftkinafnwnakvrklspfkrrkrfiklirpdansayodtlefmlasrd	300
Db	241	nlavantgilvfogftkinafnwnakvrklspfkrrkrfiklirpdansayodtlefmlasrd	300
Qy	301	fcksfwkicvhehaffrlfeepkpkpvlfrsgssfrfsgrtokqvldyvksgghkqv	360
Db	301	fcksfwkicvhehaffrlfeepkpkpvlfrsgssfrfsgrtokqvldyvksgghkqv	360
Qy	361	ferkshkhsirs1	374
Db	361	ferkshkhsirs1	374
RESULT# 3			
AA	AY91946		
ID	AAI91946	standard; Protein; 1005 AA.	
XX	XX		
XX	AA	AY91946;	
DT	19-JUL-2000	(first entry)	
XX	XX		
DE	XX	Human cytoskeleton associated protein 1 (CYSKP-1).	
XX	XX		
KW	XX	Cytoskeleton associated protein; CYSKP-1; cancer; proliferative;	
KW	XX	autoimmunity; inflammatory, vesicle trafficking; neurological;	
KW	XX	cardiovascular; cell motility; reproductive; muscle disorder.	
XX	XX		
OS	XX	Homo sapiens.	
XX	XX		
PH	XX	Key	
FT	XX	Location/Qualifiers	
FT	XX	7	
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Modified-site	15
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Modified-site	36
FT	XX	Modified-site	/note= "potential glycosylation site"
FT	XX	Modified-site	39
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Modified-site	58
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Modified-site	151
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Modified-site	159
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Modified-site	180
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Modified-site	194
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Modified-site	208
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Modified-site	212
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Modified-site	223
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Domain	225..263
FT	XX	Domain	/note= "signature sequence"
FT	XX	Domain	272..300
FT	XX	Domain	/note= "signature sequence"
FT	XX	Domain	321..367
FT	XX	Domain	/note= "signature sequence"
FT	XX	Modified-site	323
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Domain	379..408
FT	XX	Domain	/note= "signature sequence"
FT	XX	Modified-site	381
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Modified-site	402
FT	XX	Modified-site	/note= "potential phosphorylation site"
FT	XX	Domain	440..458
FT	XX	Domain	/note= "signature sequence"

FT Modified-site 449 /note= "potential phosphorylation site"

FT Modified-site 518 /note= "potential phosphorylation site"

FT Modified-site 530 /note= "potential phosphorylation site"

FT Modified-site 543 /note= "potential phosphorylation site"

FT Modified-site 544 /note= "potential phosphorylation site"

FT Modified-site 548 /note= "potential phosphorylation site"

FT Modified-site 614 /note= "potential phosphorylation site"

FT Modified-site 623 /note= "potential phosphorylation site"

FT Modified-site 647 /note= "potential phosphorylation site"

FT Modified-site 658 /note= "potential phosphorylation site"

FT Modified-site 673 /note= "potential phosphorylation site"

FT Modified-site 682 /note= "potential phosphorylation site"

FT Domain 718..721 /note= "signature sequence"

FT Modified-site 730 /note= "potential phosphorylation site"

FT Modified-site 744 /note= "potential phosphorylation site"

FT Modified-site 746 /note= "potential phosphorylation site"

FT Modified-site 748 /note= "potential phosphorylation site"

FT Modified-site 766 /note= "potential phosphorylation site"

FT Modified-site 828 /note= "potential phosphorylation site"

FT Modified-site 854 /note= "potential phosphorylation site"

FT Modified-site 879 /note= "potential phosphorylation site"

FT Modified-site 884 /note= "potential phosphorylation site"

FT Modified-site 890 /note= "potential phosphorylation site"

FT Modified-site 944 /note= "potential phosphorylation site"

FT Modified-site 952 /note= "potential phosphorylation site"

FT Modified-site /note= "potential phosphorylation site"

XX WO200017355-A2.

XX 30-MAR-2000.

XX 17-SEP-1999; 99WO-US21565.

XX 18-SEP-1998; 98US-0172226.

XX 27-APR-1999; 99US-0131321.

XX (INCY-) INCYTE PHARM INC.

XX Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;

XX Guegler KJ, Patterson C, Azimzai Y, Baughn MR;

XX WPI; 2000-283582/24.

XX N-PSDB; AAA08581.

XX Human cytoskeleton associated proteins, used to treat cell

XX proliferative, autoimmune/inflammatory, vesicle trafficking,

XX neurological, cell motility, reproductive and muscle disorders

XX Disclosure; Page 79-82; 113pp; English.

XX AAY91946-61 show human cytoskeleton associated proteins 1 to 16 (CYSKP-1

CC to CYSKP-16) respectively. The sequences can be used to treat and

CC diagnose cancer and cell proliferative, autoimmune/inflammatory, vesicle

CC trafficking, neurological, cardiovascular, cell motility, reproductive

CC and muscle disorders. Pharmaceutical compositions containing CYSKP-2 to

CC CYSKP-16 can be used to treat or prevent disorders associated with

CC decreased expression or activity of CYSKP (claimed), for example,

CC atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,

CC cancers, autoimmune/antiinflammatory disorders such as allergies, anemia,

CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,

CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma

CC and trauma. CYSKP antagonists can be used to treat or prevent a

CC disorder associated with increased expression or activity of CYSKP

CC (claimed).

XX Sequence 1005 AA;

SQ

Query Match 35.3%; Score 689.5; DB 21; Length 1005;

Best Local Similarity 42.8%; Pred. No. 3.2e-61;

Matches 148; Conservative 58; Mismatches 129; Indels 11; Gaps 6;

QY 34 TPGSKLVSIKIQMLDDTQEAPEVQAPGKVLDDAVCNHLNLVEGDYFGLFPDHHKITV 93

DB 212 tkktktvqcvktllldgteyscdlekhakgqvlfdkvcchlnllekdylfqlqespeqkn 271

QY 94 WLDLLKPIVKQIRPKHVVKFVVPDHTQLOBELTRYLFALQVKQDLAQRLLCND 153

DB 272 wldpakeikrqlnlpwft-fnvkfyppdqsteditryfclqlrqdiasgrlpcsf 330

QY 154 TSAALISHIVQSEIGDFDEALDREH---LAKNKYIP-OODALEDKIVEFHNNHIGOTP 208

DB 331 vthallgsytlqaelgdyd---peehgsidsefqaftqtkeekvaelhthrglsp 387

QY 209 AESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFGQFTKINAFNNAKVRK 268

DB 388 aqadsqflenakrlsmgyvdlhhakdsegvdklgyvcangllykdrlnrnfawpklk 447

QY 269 LSPKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFVKICVEHHAFFRLEFEPKPKPKP 328

DB 448 isykrnfyikvrpaelqfestigfklpnhraakrlwkvcehhtfyrlv-speqpoka 506

QY 329 VLFSGSSFRSGRTQKVLDYVKEGGHKKVQFERKHSKIHSIRSL 374

DB 507 kfltlgskfrysgtqastrqastltdrpaphfertsskrvs-rsl 551

RESULT 4

ID AAB95258

XX AAB95258 standard; Protein; 687 AA.

XX AC AAB95258;

XX 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:17435.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

XX 27-AUG-1999; 99JP-0300253.

XX 11-JAN-2000; 2000JP-0118776.

XX 02-MAY-2000; 2000JP-0183767.

XX 09-JUN-2000; 2000JP-0241899.





XX Disclosure; Fig 1A-B; 12pp; English.

PS A protein tyrosine phosphatase (PTPH1) (AAWL12522) catalyses the

XX dephosphorylation of proteins in which tyrosyl residues have been

CC phosphorylated through the action of a protein tyrosine kinase

CC (PTK). It localises to focal adhesions, a major site of action of

CC oncogenic PTKs. PTPH1 is the product of a cDNA clone (AAT58627)

CC obtd. from HeLa cells. The PTPH1 cDNA can be incorporated into a

CC vector (pref. retroviral) to allow expression of PTPH1 in mammalian

CC in sufficient quantities to overcome or counteract PTK activity.

CC Phosphorylation of tyrosine residues at abnormal levels is

CC prevented or reversed, resulting in the prevention or reversal of

CC malignancy of cells.

XX

XX Sequence 913 AA;

Query Match 25.9%; Score 506.5; DB 18; Length 913;

Best Local Similarity 35.0%; Pred. No. 1.4e-42;

Matches 119; Conservative 68; Mismatches 134; Indels 19; Gaps 9;

QY 27 RGQKPPPTPSGKLVSIKIQMLDDTQEAPEVQAPAGKVLDDAVCNHNLNVEGDYFGLFFP 86

Db 16 rtseelpkektseivcsihfildgvvqtkvtdqgtgvlldmvmhnhlgvtekeyfglqhd 75

QY 87 DHKKITV-WLDDLKPIVKQIRRPKHVVVFVVFPPDPHTQQLBELTRYLFALQVKODLA 145

Db 76 ddsdvsprwleaskpikrkqkgfpcctlhfrvrfipdpntlqqeqtrhyflqlkmdic 135

QY 146 QGRITCNDTSAALLISHIVQSEIGDFDEALDRE-HLAKNKYIPOQDALED---KIVEFHH 201

Db 136 egritcplnsavvlasyavqshgfygdnssihphgylsdshfipdqn--edfltkveslhe 193

QY 202 NHIGQTPAESDFQLLEIARLEMYGIRLHPAKDREGCKINLAVANTGILVFGQTKINAF 261

Db 194 qhsglkqseascyiniartldfygvelhsgrdlhnldlmigiasagvavrykictsfy 253

QY 262 NNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSPWVKTCVEHAFRLEFEE 321

Db 194 qhsglkqseascyiniartldfygvelhsgrdlhnldlmigiasagvavrykictsfy 253

RESULT 6

AAV25156

ID AAY25156 standard; Protein; 913 AA.

XX AAY25156;

XX 07-SEP-1999 (first entry)

XX Human PTPH1 protein.

XX PTPH1; human; protein tyrosine phosphatase; focal adhesion; cancer;

XX localisation; treatment; overexpression; oncogenic; cell transformation;

XX prevention; phosphotyrosine; disease; malignant.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Region 30..357

XX /note- "region of homology to the N-terminal domain

XX of band 4.1, ezrin and talin. This region is

XX known to be important for localisation to focal

XX adhesions"

XX Modified-site 372

XX /note- "potential p34cdc2 phosphorylation site"

XX Modified-site 381

XX /note- "potential p34cdc2 phosphorylation site"

FT Modified-site 424..428

FT /note- "potential casein kinase II phosphorylation site"

FT Modified-site 434

FT /note- "potential p34cdc2 phosphorylation site"

FT Modified-site 438..442

FT /note- "potential casein kinase II phosphorylation site"

FT Modified-site 489..492

FT /note- "potential casein kinase II phosphorylation site"

FT Modified-site 514..518

FT /note- "potential casein kinase II phosphorylation site"

FT Modified-site 543..547

FT /note- "potential casein kinase II phosphorylation site"

FT Modified-site 607..610

FT /note- "potential casein kinase II phosphorylation site"

XX

XX US5863781-A.

XX

XX 26-JAN-1999.

XX

XX 04-DEC-1996; 96US-0759536.

XX

XX 01-MAR-1991; 91US-0663579.

XX 14-MAR-1990; 90US-0494036.

XX 16-AUG-1993; 93US-0107420.

XX 04-DEC-1996; 96US-0759536.

XX

XX (COLD-) COLD SPRING HARBOR LAB.

XX

XX Tonks NK;

XX

XX WPI; 1999-131308/11.

XX N-PSDB; AAX78463.

XX

XX Protein tyrosine phosphatase PTPH1 - encoded by DNA of HeLa cells

XX

XX Claim 1; Fig 1A-B; 12pp; English.

XX

XX This sequence represents a novel protein tyrosine phosphatase, PTPH1,

XX isolated from HeLa cells. The protein of the invention appears to

XX localise to focal adhesions and is therefore potentially useful in the

XX treatment of cancer. Overexpression of PTPH1 can be used to counter the

XX effects of oncogenic protein tyrosine kinases such as those of

XX transforming viruses and for interfering with or reversing cell

XX transformation. This would provide a means of preventing or reversing

XX abnormally high levels of phosphotyrosine associated with any disease or

XX condition such as preventing or reversing malignancy associated with the

XX activity of a protein tyrosine kinase.

XX

XX Sequence 913 AA;

Query Match 25.9%; Score 506.5; DB 20; Length 913;

Best Local Similarity 35.0%; Pred. No. 1.4e-42;

Matches 119; Conservative 68; Mismatches 134; Indels 19; Gaps 9;

QY 27 RGQKPPPTPSGKLVSIKIQMLDDTQEAPEVQAPAGKVLDDAVCNHNLNVEGDYFGLFFP 86

Db 16 rtseelpkektseivcsihfildgvvqtkvtdqgtgvlldmvmhnhlgvtekeyfglqhd 75

QY 87 DHKKITV-WLDDLKPIVKQIRRPKHVVVFVVFPPDPHTQQLBELTRYLFALQVKODLA 145

Db 76 ddsdvsprwleaskpikrkqkgfpcctlhfrvrfipdpntlqqeqtrhyflqlkmdic 135

QY 146 QGRITCNDTSAALLISHIVQSEIGDFDEALDRE-HLAKNKYIPOQDALED---KIVEFHH 201

Db 136 egritcplnsavvlasyavqshgfygdnssihphgylsdshfipdqn--edfltkveslhe 193

QY 202 NHIGQTPAESDFQLLEIARLEMYGIRLHPAKDREGCKINLAVANTGILVFGQTKINAF 261

Db 194 qhsglkqseascyiniartldfygvelhsgrdlhnldlmigiasagvavrykictsfy 253

QY 262 NNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSPWVKTCVEHAFRLEFEE 321

XX



CC system disorders, muscular disorders, reproductive disorders,  
 CC gastrointestinal disorders, wounds, renal disorders, infectious  
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and  
 CC AAB54007 represent sequences used in the exemplification of the present  
 CC invention.

SQ Sequence 635 AA;

Query Match 18.4%; Score 359; DB 21; Length 635;  
 Best Local Similarity 27.9%; Pred. No. 9.9e-28;  
 Matches 107; Conservative 71; Mismatches 148; Indels 58; Gaps 13;

QY 7 RPTPGSRLGAPEN-----GISTLERGQPPPTPSGKLVSIKIOMLDDTQAEFVPPQR 59  
 Db 19 rstagaglgqlrccwivvfgiedthqkpkmp-----kpinrvvtmdaele-faiqpn 72

QY 60 APGKVLDAVNCNHLNVEGDFYGLFPPDHKKITWLDLLKPI-VKQIRPKHVHVVFVK 118  
 Db 73 ttgqlfdqvvktiglrrevvyfghyvdnkgfptwklklkksaqevrkenplqfkfrak 132

QY 119 FFPDP-HTOLQEBELRYLFAOVKQDLAQGLTCNDTSAALLSHIVQSEIGDFDEALDR 177  
 Db 133 fypedvaeelqgdtqklflqkvegilsdeilycpptavllgsyavqakfgdynkevkh 192

QY 178 E-HLAKNKYIPQ-----QDALEKIVFEFHHNHIGQTPAESDFOLLEIARRLEMYGI 227  
 Db 193 sgyllserlipqvmqdkhitrqwedriqwhaehrgmlkdnamleylkiadlemygi 252

QY 228 RLHPAKDREGTKINLAVANTGIILVFGFTKIN---AFNNAKVRKLSFKRRFLI----KL 280  
 Db 253 nyfeiknkgtdlwlvgdaigltnyekddkltpkigfpwsefnisfndkfkvikpidkk 312

QY 281 RPANSAYQDTLEFLMASRDFCKSFWKICVCEHHAFFELFEEPKPKPKPVLFSGSFRFS 340  
 Db 313 apd-----fvyaprlrinkrlqcmgnhely-mrrrkpdktiev----- 351

QY 341 GRTQKQVLDVYKGGHKKVQFERK 364  
 Db 352 ---qgmkaqaeehkhq-qlerq 370

RESULT 9  
 AAW67438  
 ID AAW67438 standard; Protein; 1174 AA.  
 XX AAW67438;  
 AC AAW67438;  
 XX AAW67438;  
 DT 02-MAR-1999 (first entry)  
 DE Human protein tyrosine phosphatase D1 protein.  
 XX Human; protein tyrosine phosphatase; skeletal muscle; hybridisation;  
 KW glycoprotein; probe.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX Key Location/Qualifiers  
 FH 24..339  
 FT Domain /note="ezrin-like domain"  
 FT Binding-site 158..161  
 FT Binding-site 207..210 /note="SH2 binding motif"  
 FT Binding-site 217..220 /note="SH2 binding motif"  
 FT Binding-site 217..220 /note="SH2 binding motif"  
 FT Modified-site 336..338 /note="SH2 binding motif"  
 FT Modified-site 354..356 /note="PYX motif, putative site of Tyr phosphorylation"  
 FT Modified-site 408..410 /note="PYX motif, putative site of Tyr phosphorylation"  
 FT Modified-site 523..525 /note="PYX motif, putative site of Tyr phosphorylation"

FT Modified-site /note="PYX motif, putative site of Tyr phosphorylation" 525..527  
 FT Binding-site /note="PYX motif, putative site of Tyr phosphorylation" 565..574  
 FT Modified-site /note="SH3 binding domain" 568..570  
 FT Region /note="PYX motif, putative site of Tyr phosphorylation" 712..722  
 FT Domain /note="acidic stretch" 921..1166  
 FT /note="protein tyrosine phosphatase domain" 921..1166  
 XX US5831009-A.  
 XX 03-NOV-1998.  
 XX 22-MAY-1995; 95US-0446345.  
 XX 27-APR-1994; 94US-0234440.  
 PR 15-AUG-1992; 92US-0923740.  
 PR 22-MAY-1995; 95US-0446345.  
 XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
 PA Moller KB, Moller NPH, Ullrich A;  
 PI WPI: 1998-609316/51.  
 DR N-PSDB; AAV34368.  
 XX Protein tyrosine phosphatase D1 - useful in screening assays for  
 PT e.g. agonists or antagonists  
 XX Claim 1; Fig 5A-J; 61pp; English.  
 XX This is the sequence of the complete human protein tyrosine phosphatase  
 CC D1 (PTP-D1). The coding sequence was isolated from a human skeletal  
 CC muscle cDNA library. The phosphatase is used e.g. in a hybridisation  
 CC assay for detecting a nucleic acid encoding a normal or mutant PTP-D  
 CC protein or glycoprotein in a cell or subject, using a nucleic acid  
 CC comprising at least part of a normal or mutant PTP-D protein or  
 CC glycoprotein as a probe; or a screening assay for a compound capable  
 CC of binding to a PTP-D protein or glycoprotein, using a PTP-D protein  
 CC or glycoprotein or the compound-binding portion of it attached to a  
 CC solid support.

SQ Sequence 1174 AA;  
 Query Match 18.2%; Score 356; DB 19; Length 1174;  
 Best Local Similarity 27.1%; Pred. No. 5e-27;  
 Matches 94; Conservative 70; Mismatches 147; Indels 36; Gaps 8;

QY 25 LERQKPPPTPSGKLVSIKIOMLDDTQAEFVPPQRAPGKVLDAVNCNHLNVEGDFGLE 84  
 Db 9 lkrtrrytvssksclva-riqlinnfevtlsvestqgesleavagrlreivtyfslw 67

QY 85 FPDHKKITVWLDLLKPIVKQI-RRPKHVHVVFVFPDPDHTOLOEBELRYLFAOVKQD 143  
 Db 68 yynkqqrtrwvdeklpkqlkqdyaleptvyfgyvsvsqqlqgeitryqylqlkdd 127

QY 144 LAQGLTCNDTSAALLSHIVQSEIGDFDEALDREHLAKNKYIP-----QQDALED---K 195  
 Db 128 ilegsipctleqaiqlaglavqadfgdfqyesqdfqlqfalfpvgwlgdekveeatqk 187

QY 196 IVEFHHNHIGQTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIILV-FQG 254  
 Db 188 vallhdkyrgltapdaemlymgevermdgygeesypakdsqgsdisagaclegifvkhkn 247

QY 255 FTKINAFNNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHA 314  
 Db 248 grhpvvfrwhdianmshnksffalel---ank---ectiqfqtmedmetakiyrlvarhk 302

QY 315 FFLRLE-----EPKPKPKPVLFSGSFRFSG 341





CC neuromas) of the eighth cranial nerve. Other symptoms of NF2 include  
 CC cranial meningiomas, spinal nerve root schwannomas and presenile lens  
 CC opacities. The NF2 gene, mapped to chromosomal region 22q12 between the  
 CC loci D22S1 and D22S28, acts as a tumour suppressor. The NF2 gene is  
 CC alternatively spliced resulting in three different isoforms encoding  
 CC three different proteins, merlin isoforms I-III, which are likely to  
 CC have differing functions. Merlin stands for moesin-erzin-radixin like  
 CC protein, so called due to substantial homology with these three  
 CC proteins. NF2 isoform II cDNA and protein are useful in diagnosing NF2  
 CC disease. Merlin protein products act as tumour suppressors and can be  
 CC used to suppress tumour growth, as can the cDNA sequence in gene therapy  
 CC applications. Antibodies raised against merlin proteins are useful as  
 CC tumour targeting agents. The human NF2 isoform was identified in a  
 CC foetal mouse brain cDNA library.

XX Sequence 591 AA;

Query Match 16.8%; Score 328; DB 18; Length 591;

Best Local Similarity 27.2%; Pred. No. 1.3e-24;

Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;

QY 23 STLERGQPPPTPSGKLVSIKIQMLDDTQEAPEVQAPGKVLDDAVCNHNLNVEGDYFG 82

DB 12 ssLkr--kqpkftvriutm-----daemefncemkkgkldflvrtlgretwffg 63

QY 83 LEFPDHKKITVWLDLLKPIV-KQIRPKHVVKVVFVPPDP-HTQLOEELTRYLFALQV 140

DB 64 lqy-tikdtvawlkmdkvdhdskeepvtfhflakfyenaeeelvqeiltghiflqv 122

QY 141 KQDLAQGRITCNDTSAALLISHIVQSEIGDFDEAL-DREHLAKNKYIPQO-----D 190

DB 123 kkgildexklycppeasvllasyavqakgydpsvvhkrgflaqeelpkrvinlygmtpe 182

QY 191 ALEDKIVEFHNNHIGQTPAESDFQLELTARLEMGIRLHPAKDREGTKINLAVANTGTL 250

DB 183 mweeritawyaehrgradeaemeylkiaqdlmgyvnyfairnkkgtellvgdalgih 242

QY 251 VQGGTTKIN---AFNWKVKRLSEKRRFLIKLRPDANSAYODTLEFLMASRDFCKSPWK 307

DB 243 lydpentpklspfwneirnisysdkeftik----pldkkidvfkfnssklrvnkliliq 298

QY 308 ICVEHHAF 316

DB 299 lcighndlf 307

RESULT 14

ID AAR60398

XX AAR60398 standard; Protein; 595 AA.

AC AAR60398;

XX 20-APR-1995 (first entry)

DE Merlin protein encoded by cDNA clone JJR-1.

XX Polymerase chain reaction; PCR; amplify; primer; bi-lateral schwannoma;  
 KW sequence-tagged site assay; chromosome 22; NF2; deletion; hearing loss;  
 KW neurofibromatosis; merlin; moesin-erzin-radixin-like protein; D22S28;  
 KW tumour suppressor; activity; meningioma; cytoskeleton; gene therapy;  
 KW merlin-associated tumour; D22S1; posterior capsular lens opacity;  
 KW deafness; balance disorder; paralysis; ss.

XX Homo sapiens.

PN EP613945-A.

XX 07-SEP-1994.

XX 25-FEB-1994; 94EP-0301367.

XX 25-FEB-1993; 93US-0022034.

PR 04-MAR-1993; 93US-0026063.

PR 19-AUG-1993; 93US-0108808.

XX 22-DEC-1993; 93US-0171718.

PA (GEOH) GEN HOSPITAL CORP.

PI Gusella JF, MacCollin MM, Trofatter JA;

XX WPI; 1994-272992/34.

DR N-PSDB; AAQ71077.

XX The tumour suppressor gene merlin - for treatment and diagnosis

PT of tumours and neurofibromatosis (NF2)

XX Example 3; Page 35-38; 86pp; English.

CC This sequence is encoded by the clone JJR-1 and represents the merlin  
 CC protein. The cDNA sequence contains eight cloned exons sequences and  
 CC encodes this protein of 69 kD. The cDNA contains no evidence of a  
 CC poly-A tail and spans at least 50 kb of genomic DNA. NF2 is a neuro-  
 CC fibromatosis which is characterised by bi-lateral schwannomas. The  
 CC NF2 "gene" has been shown by linkage studies to be assigned to  
 CC chromosome 22. The missing or mutated gene in NF2 patients has been  
 CC shown to be the merlin gene. The gene encodes a protein, merlin  
 CC (moesin-erzin-radixin-like protein), which possesses tumour suppressor  
 CC activity, and whose tumour suppressor activity is mediated by inter-  
 CC actions with the cytoskeleton. The merlin gene is found on chromosome  
 CC 22 between the known markers D22S1 and D22S28. In patients suffering  
 CC from NF2, the merlin gene is either lost or mutated. A mutant merlin  
 CC protein may be encoded by a gene in which a mutation of A to T at the  
 CC first position of the codon encoding amino acid 220 causes the  
 CC substitution of Tyr for Asn. The merlin gene may be used in gene  
 CC therapy for the treatment of a merlin-associated tumour or NF2, or  
 CC for prevention of schwannoma, meningioma, posterior capsular lens  
 CC opacities, deafness or hearing loss, balance disorders or paralysis.

XX Sequence 595 AA;

Query Match 16.8%; Score 328; DB 15; Length 595;

Best Local Similarity 27.2%; Pred. No. 1.3e-24;

Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;

QY 23 STLERGQPPPTPSGKLVSIKIQMLDDTQEAPEVQAPGKVLDDAVCNHNLNVEGDYFG 82

DB 12 ssLkr--kqpkftvriutm-----daemefncemkkgkldflvrtlgretwffg 63

QY 83 LEFPDHKKITVWLDLLKPIV-KQIRPKHVVKVVFVPPDP-HTQLOEELTRYLFALQV 140

DB 64 lqy-tikdtvawlkmdkvdhdskeepvtfhflakfyenaeeelvqeiltghiflqv 122

QY 141 KQDLAQGRITCNDTSAALLISHIVQSEIGDFDEAL-DREHLAKNKYIPQO-----D 190

DB 123 kkgildexklycppeasvllasyavqakgydpsvvhkrgflaqeelpkrvinlygmtpe 182

QY 191 ALEDKIVEFHNNHIGQTPAESDFQLELTARLEMGIRLHPAKDREGTKINLAVANTGTL 250

DB 183 mweeritawyaehrgradeaemeylkiaqdlmgyvnyfairnkkgtellvgdalgih 242

QY 251 VQGGTTKIN---AFNWKVKRLSEKRRFLIKLRPDANSAYODTLEFLMASRDFCKSPWK 307

DB 243 lydpentpklspfwneirnisysdkeftik----pldkkidvfkfnssklrvnkliliq 298

QY 308 ICVEHHAF 316

DB 299 lcighndlf 307

RESULT 15

AAW94459

ID AAW94459 standard; Protein; 596 AA.

XX AC AAW94459;

Job time: 334 sec

```

XX 16-APR-1999 (first entry)
XX Human neurofibromatosis type 2 transcript isoform I protein.
DE
XX
XX Human neurofibromatosis type 2; NF2; tumour suppressor; cancer;
KW diagnosis; gene therapy.
XX
OS Homo sapiens.
XX
PN US5872214-A.
XX
PD 16-FEB-1999.
XX
XX 04-APR-1996; 96US-0628145.
XX
PR 10-JAN-1994; 94US-0179738.
PR 04-APR-1996; 96US-0628145.
XX
XX (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
XX Bianchi AB, Kley NA, Seizinger BR;
XX
DR WPI; 1999-166715/14.
XX
PT Proteins from neurofibromatosis type 2 transcript isoforms - used
PT for diagnosis or inhibition of tumours, and generation of antibodies
XX
PS Example 1; Fig 1; 45pp; English.
XX
CC The present sequence represents human neurofibromatosis type 2 (NF2)
CC transcript isoform I. NF2 polynucleotides can be used for diagnosing
CC NF2 diseases, for inhibiting growth of tumours associated with NF2
CC mutations (including expression from cDNA introduced in gene therapy
CC vectors) and to raise antibodies (useful as tumour targeting agents,
CC since specific isoforms are often tumour-specific) and as immunoassay
CC reagents for detecting NF2-expression products. NF2 is a tumour
CC suppressor protein, and so has anticancer activity.
XX
SQ Sequence 596 AA;

Query Match 16.8%; Score 328; DB 20; Length 596;
Best Local Similarity 27.2%; Pred. No. 1.3e-24;
Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;

QY 23 STLERQKPPPTPSGKLVSIKQLMDTQEAPEVQAPAGKVLDDAVCNHNLNVEGDYFG 82
DB 12 ssikr--kqpkftftrivlm-----daemefncemkwgkdlfdlvcrtlgltretwffg 63
QY 83 LEFPDHKKITVMDLLKPIV-KQIRPKHWVVKVVFPPD-HTOLOBELTRYLFALOV 140
DB 64 lgy-tikdtvawlkmdkvidhdvskeepvtfhflakfypenaeelqvqitqhlfflgy 122
QY 141 KQDLAQGR LTCNDTSAALLSHIVQSEIGDFEAL-DREHLAKNKYIPQO-----D 190
DB 123 kqildekicycpeasvllasyavqakgydpsvhkgrflaqeelpkrvlnlyqmtpe 182
QY 191 ALEDKIVEFHNNHIGTPAESFQLEIARLEMYGIRLHPAKDREGTKINLAVANTGIL 250
DB 183 mweiritavyaehrgardeaemeykiaqdiemygvnyfairnkkgtelllgvdalgh 242
QY 251 VFQGFTKIN--AFNNAKVKRLSFKKRFLIKLRPDANSAYODTLEFLMASRDFCKSFWK 307
DB 243 iydpenrltpklsfpwneirnisysdkeftik----pldkkidvfkfnsskkrvnklliq 298
QY 308 ICVEHHAFF 316
DB 299 lcignhldf 307
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:55:08 ; Search time 36.72 Seconds  
(without alignments)  
229.200 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_1\_374

Perfect score: 1956

Sequence: 1 MGIEQPTFGSLGAPENS.....GHKKVQFERKSHKHSIRSL 374

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	646	33.0	345	2	US-08-446-345-40
2	612	31.3	288	3	US-09-100-804-14
3	522	26.7	247	3	US-09-100-804-13
4	500.5	25.6	288	3	US-09-100-804-15
5	499	25.5	342	2	US-08-446-345-39
6	356	18.2	1174	2	US-08-446-345-36
7	344	17.6	296	3	US-09-100-804-12
8	328	16.8	591	1	US-08-179-738-10
9	328	16.8	591	2	US-08-628-145-10
10	328	16.8	595	1	US-08-171-718-16
11	328	16.8	595	3	US-08-478-087-16
12	328	16.8	596	1	US-08-179-738-3
13	328	16.8	596	2	US-08-628-145-3
14	325	16.6	584	1	US-08-179-738-7
15	325	16.6	584	2	US-08-628-145-7
16	323	16.5	591	1	US-08-179-738-5
17	323	16.5	591	2	US-08-628-145-5
18	323	16.5	596	1	US-08-179-738-2
19	323	16.5	596	2	US-08-628-145-2
20	265	13.5	303	3	US-09-100-804-11
21	265	13.5	2465	3	US-08-596-291-3
22	265	13.5	2465	3	US-09-100-804-3
23	265	13.5	2466	3	US-09-080-855-12
24	265	13.5	2466	5	PCT-US94-09943-2
25	265	13.5	2485	4	US-09-290-640-46
26	91	4.7	522	1	US-08-625-322-2
27	89.5	4.6	212	4	US-09-259-109-2

28	87	4.4	1164	4	US-08-923-992A-10	Sequence 10, Appl
29	86	4.4	472	1	US-08-203-905B-14	Sequence 14, Appl
30	85.5	4.4	1704	4	US-08-485-355B-40	Sequence 40, Appl
31	84.5	4.3	1462	3	US-07-792-600-31	Sequence 31, Appl
32	84.5	4.3	1462	3	US-09-157-021-31	Sequence 31, Appl
33	84.5	4.3	1462	3	US-09-156-842-31	Sequence 31, Appl
34	84	4.3	984	1	US-08-242-932-2	Sequence 2, Appl
35	84	4.3	984	1	US-08-714-481-2	Sequence 2, Appl
36	84	4.3	984	5	PCT-US95-06111-2	Sequence 2, Appl
37	84	4.3	1164	4	US-08-923-992A-2	Sequence 2, Appl
38	83.5	4.3	633	2	US-08-736-770-3	Sequence 3, Appl
39	82	4.2	1128	4	US-08-923-992A-6	Sequence 6, Appl
40	82	4.2	1513	5	PCT-US93-03076-2	Sequence 2, Appl
41	81.5	4.2	517	2	US-08-967-508-19	Sequence 19, Appl
42	81.5	4.2	517	3	US-08-967-506-19	Sequence 19, Appl
43	81.5	4.2	517	5	PCT-US94-02552-19	Sequence 19, Appl
44	81.5	4.2	559	2	US-08-967-508-9	Sequence 9, Appl
45	81.5	4.2	559	3	US-08-967-506-9	Sequence 9, Appl

#### ALIGNMENTS

RESULT 1  
US-08-446-345-40  
; Sequence 40, Application US/08446345  
; Patent No. 5831009  
; GENERAL INFORMATION:  
; APPLICANT: Ullrich, Axel  
; APPLICANT: Moller, Niels P.H.  
; APPLICANT: Moller, Karin B.  
; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE  
; TITLE OF INVENTION: PHOSPHATASES PTP-D1  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: N.Y.  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,345  
; FILING DATE: 22-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/234,440  
; FILING DATE: 28-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30742  
; REFERENCE/DOCKET NUMBER: 7683-054  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 345 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-446-345-40

Query Match 33.0%; Score 646; DB 2; Length 345;  
Best Local Similarity 40.9%; Pred. No. 4.5e-59;





```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,345
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/234,440
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-054
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212) 869-8864
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 342 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-446-345-39

```

Query Match 25.5%; Score 499; DB 2; Length 342;

Best Local Similarity 34.7%; Pred. No. 9,le-44;

Matches 114; Conservative 66; Mismatches 133; Indels 16; Gaps 7;

```

QY 27 RGQKPPPTSGKLVSTIKIQLDDTOEAFVPPQAPGKVLDDAVCNHLNVLVEGDYGLLEPP 86
DB 16 RISELPEKTRSEVICSIFHVGVOVTFKVTQDGTQGVLDWMVHNLGVTEKEYFGLQHD 75
QY 87 DHKKITV-WLDDLLKPIVKQIRRPKHVVVKKVFFPPDPDHTLQEEELTRYLFALQVKQDILA 145
DB 76 DSDVDSPRLEASLPKQKGGFPCTLHFRVRFPIPDNTLQEQTRHLYFLQKMDIC 135
QY 146 QGRLCNTDTSALLSHIVQSIGDFDEALDRE-HLAKNKYIPQOQDALED---KIVEFH 201
DB 136 EGRLTCLPLNSAVLASYAFVSHFGDYNSSIHHPGYSLSHFPDQD--EDFLTQVESLHE 193
QY 202 NHIGQTPAESDFOLLEIARRLEMVIGIRLHPAKDREGTKINLAVANTGILVFOGFTKINAF 261
DB 194 QHSGLKQSAESCYNIARTLDFYGVELHSGRDLNLDLMIGIASAGVAYRYKICTSY 253
QY 262 NNAKVKLSFKRRRLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVVEHFAFFLPEE 321
DB 254 PWNILKISFKRKKFFIHQRQKQAESREHIVAFNMLNYSRCKNLWKSCVEHHTFFQA-KK 312
QY 322 PKPKPKPVL---FSRGSFRFSGRTQKV 347
DB 313 LLPQEKNVLSQYWTGSR-----NTKKS 336

```

RESULT 6

US-08-446-345-36

; Sequence 36, Application US/08446345

; Patent No. 5831009

; GENERAL INFORMATION:

; APPLICANT: Ullrich, Axel

; APPLICANT: Moller, Niels P.H.

; APPLICANT: Moller, Karin B.

; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE

; TITLE OF INVENTION: PHOSPHATASES PTP-DI

; NUMBER OF SEQUENCES: 41

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds

; STREET: 1155 Avenue of the Americas

; CITY: New York

```

; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,345
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/234,440
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-054
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1174 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-446-345-36

```

Query Match 18.2%; Score 356; DB 2; Length 1174;

Best Local Similarity 27.1%; Pred. No. 4.7e-28;

Matches 94; Conservative 70; Mismatches 147; Indels 36; Gaps 8;

```

QY 25 LERGQPPPTSGKLVSTIKIQLDDTOEAFVPPQAPGKVLDDAVCNHLNVLVEGDYGLLE 84
DB 9 LKTRRYTVSSKSLVA-RIQLLNNEFVEFTLSVESTGOESLEAQAQRLREVTYFSLW 67
QY 85 FPDHKKITVWLDLLKPIVKQI-RRPKHVVVKKVFFPPDPDHTLQEEELTRYLFALQVKOD 143
DB 68 YYNQQRNRWDLERPLKKQLDKYALEPTVYGVVFPVPSVQLQOQEIETRYQYVQLQKKD 127
QY 144 LAQGRLTCDTSAALLSHIVQSIGDFDEALDREHLAKNKYIP-----QOQDALED---K 195
DB 128 ILEGSIPTCLEQAIQAGLAVQADFGDFQYESQDFLQKFAFPVGNLQDEKVLLEATQK 187
QY 196 IYEFHHNHIGQTPAESDFOLLEIARRLEMVIGIRLHPAKDREGTKINLAVANTGILV-FQG 254
DB 188 VALLHQKYRGLTAPDAEMLYMQEVERMDGYGESVPKDSQSGDISIGACLEGIFVKHKN 247
QY 255 FTKINAFNWKVKLSFKRRRLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVVEHHA 314
DB 248 GRHPVYFRWHDIANMHNKSNFFALEL---ANK--EETIQFQTEDEMTAKYIWRLCVARHK 302
QY 315 FPRLFE-----EPKPKPKVLFSGSFRFSG 341
DB 303 FYRLNQCNIQTOTVTVPNPIRRSSSRMSLPPKQPYVMPPPPQLHYNG 349

```

RESULT 7

US-09-100-804-12

; Sequence 12, Application US/09100804

; Patent No. 6066472

; GENERAL INFORMATION:

; APPLICANT: GONEZ, LEONEL JORGE

; APPLICANT: SARAS, JAN

; APPLICANT: CLAESSON-WELSH, LENA

; APPLICANT: HELDIN, CARL-HENRIK

; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL

; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN



[illegible][illegible]

Db 123 KKQILDEKIYCPPEASVLLASVAVQAKYGDYDPSVHKRGFLAQEBELLKRVINLYQMTPE 182  
Qy 191 ALEDKIVEFHNNHIGQTPAESDFOLLEIARLEMVGIKRLHPAKDREGTKINLAVANTGIL 250  
Db 183 MWEERITAYAEHRGRARDEAEMEYLKIAQDLEMVGVNYFAIRNKKGTTELLIGVDALGLH 242  
Qy 251 VFOGFTKIN--AFNNAVKRLSPKRRFLIKLRDPDANSAYQDTLEFLMASRDFCKSEWK 307  
Db 243 IYDPENRLTPKISFPWNEIRNISYSKDEFTIK-----PLDKKIDVFKFNSSKLRVKNKLIQ 298  
Qy 308 ICVEHHAFF 316  
Db 299 LCIGNHDLF 307  
RESULT 11  
US-08-478-087-16  
; Sequence 16, Application US/08478087  
; Patent No. 6077685  
; GENERAL INFORMATION:  
; APPLICANT: Trofatter, James A.  
; APPLICANT: MacCollin, Mia M.  
; APPLICANT: Gusella, James F.  
; TITLE OF INVENTION: Tumor Suppressor Gene Merlin and Uses  
; TITLE OF INVENTION: Thereof  
; NUMBER OF SEQUENCES: 120  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox  
; STREET: 1100 New York Avenue, N.W., Suite 600  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-3934  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478.087  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/171,718  
; FILING DATE: 22-DEC-1993  
; APPLICATION NUMBER: US 08/108,808  
; FILING DATE: 19-AUG-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/022,034  
; FILING DATE: 25-FEB-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/026,063  
; FILING DATE: 04-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Brown, Anne  
; REGISTRATION NUMBER: 36,463  
; REFERENCE/DOCKET NUMBER: 0609.3850003  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 371-2600  
; TELEFAX: (202) 371-2540  
; INFORMATION FOR SEQ ID NO: 16:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 595 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-478-087-16

Query Match 16.8%; Score 328; DB 3; Length 595;  
Best Local Similarity 27.2%; Pred. No. 1.4e-25;  
Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;

Qy 23 STLERGQKPPPTPSCKLYSIKIQMLDDTQEAPEVQORAPGKVLDDAVCNHLNLVSGDYEG 82  
Db 12 SSLKR--KOPKTFTRIIVTM-----DAEMEFNCMKWKGDLDVLVCTLGLRETWFFG 63  
Qy 83 LEFPDHKKTIVWLDLKPIV-KQIRRPKHVVVVKFVVKFPDPD-HTQLOBELTRYLFALQV 140  
Db 64 LOY-TIKDTVAVLWKMDKKVLDHDSKEEPVTFHFLAKYPYPENAEELVQEIQHLEFFQV 122  
Qy 141 KODLAQGLTCDNTSAALLISHIVQSEIGDFDEAL-DREHLAKNKYIPQ-----D 190  
Db 123 KKQILDEKIYCPPEASVLLASVAVQAKYGDYDPSVHKRGFLAQEBELLKRVINLYQMTPE 182  
Qy 191 ALEDKIVEFHNNHIGQTPAESDFOLLEIARLEMVGIKRLHPAKDREGTKINLAVANTGIL 250  
Db 183 MWEERITAYAEHRGRARDEAEMEYLKIAQDLEMVGVNYFAIRNKKGTTELLIGVDALGLH 242  
Qy 251 VFOGFTKIN--AFNNAVKRLSPKRRFLIKLRDPDANSAYQDTLEFLMASRDFCKSEWK 307  
Db 243 IYDPENRLTPKISFPWNEIRNISYSKDEFTIK-----PLDKKIDVFKFNSSKLRVKNKLIQ 298  
Qy 308 ICVEHHAFF 316  
Db 299 LCIGNHDLF 307  
RESULT 12  
US-08-179-738-3  
; Sequence 3, Application US/08179738  
; Patent No. 5578462  
; GENERAL INFORMATION:  
; APPLICANT: Seizinger, Bernd R.  
; APPLICANT: Kley, Nikolai A.  
; APPLICANT: Bianchi, Albert B.  
; TITLE OF INVENTION: No. 5578462el NF2 Isoforms  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Reed & Robins  
; STREET: 635 Bryant Street  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: U.S.A  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/179,738  
; FILING DATE: 10-JAN-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Robins, Roberta L.  
; REGISTRATION NUMBER: 33,208  
; REFERENCE/DOCKET NUMBER: 5998-0017  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 617-8999  
; TELEFAX: (415) 327-3231  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 596 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapien  
US-08-179-738-3

Query Match 16.8%; Score 328; DB 1; Length 596;  
Best Local Similarity 27.2%; Pred. No. 1.4e-25;  
Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;

QY 23 STLERGQKPPPTSGKLVSIIKIOMLDDTQEAPEVFORAPGKVLDDAVCNHNLNVEGDYFG 82  
 Db 12 SSLKR--KQPKTFTVRITVM-----DAEMEFNCMKWKGKDLFDLVCRITGLRETWTFG 63  
 QY 83 LEFPDHKKITVMDLLKPIV-KQIRPKHVVVVVFVVFPPD-HTQLOBELTRYLFALOV 140  
 Db 64 LOY-TIKDVTAVLMDKVDLHDVSKKEPVTFHFLAKFYPENAEELVQETQHLEFFLOV 122  
 QY 141 KDLAQGLTCDNTSAALLISHIVQSEIGDFDEAL-DREHLAKNKYIPOQ-----D 190  
 Db 123 KQILDEKICYCPPEASVLLASVAVQAKYGDYDPSVHKRGFLAQEELLPRKVINLYQMTPE 182  
 QY 191 ALEDKIVEPHHHIGOTPAESDFOLLEIARLEMYGIRLHPAKDRGKTINLAVANTGIL 250  
 Db 183 MWEERITANYAEHRGRARDEAEWEYLKIAQDLEMYGVNYFAIRNKKGTLLGLVDALGLH 242  
 QY 251 VFQGGTKIN---AFNWKVKLSFKRKLRLKRPDANSAYODTLEFLMASRDFCKSFVK 307  
 Db 243 IYDPENRLPKISFPWNEIRNISYSDEFTIK-----PLDKKIDVFKFNSSKLRVNKLILQ 298  
 QY 308 ICVEHHAFF 316  
 Db 299 LCIGNHDLF 307

RESULT 13  
 US-08-628-145-3  
 ; Sequence 3, Application US/08628145  
 ; Patent No. 5872214  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Seizinger, Bernd R.  
 ; APPLICANT: Kley, Nikolai A.  
 ; APPLICANT: Bianchi, Albert B.  
 ; TITLE OF INVENTION: No. 5872214el NF2 Isoforms  
 ; NUMBER OF SEQUENCES: 26  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Reed & Robins  
 ; STREET: 635 Bryant Street  
 ; CITY: Palo Alto  
 ; STATE: California  
 ; COUNTRY: U.S.A  
 ; ZIP: 94301  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/628,145  
 ; FILING DATE: 04-APR-1996  
 ; CLASSIFICATION: 530  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/179,738  
 ; FILING DATE: 10-JAN-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Robins, Roberta L.  
 ; REGISTRATION NUMBER: 33,208  
 ; REFERENCE/DOCKET NUMBER: 5998-0017  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 617-8999  
 ; TELEFAX: (415) 327-3231  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 596 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; FRAGMENT TYPE: N-terminal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Homo sapien  
 ; US-08-628-145-3

Query Match 16.8%; Score 328; DB 2; Length 596;  
 Best Local Similarity 27.2%; Pred. No. 1.4e-25;  
 Matches 84; Conservative 77; Mismatches 120; Indels 28; Gaps 9;  
 QY 23 STLERGQKPPPTSGKLVSIIKIOMLDDTQEAPEVFORAPGKVLDDAVCNHNLNVEGDYFG 82  
 Db 12 SSLKR--KQPKTFTVRITVM-----DAEMEFNCMKWKGKDLFDLVCRITGLRETWTFG 63  
 QY 83 LEFPDHKKITVMDLLKPIV-KQIRPKHVVVVVFVVFPPD-HTQLOBELTRYLFALOV 140  
 Db 64 LOY-TIKDVTAVLMDKVDLHDVSKKEPVTFHFLAKFYPENAEELVQETQHLEFFLOV 122  
 QY 141 KDLAQGLTCDNTSAALLISHIVQSEIGDFDEAL-DREHLAKNKYIPOQ-----D 190  
 Db 123 KQILDEKICYCPPEASVLLASVAVQAKYGDYDPSVHKRGFLAQEELLPRKVINLYQMTPE 182  
 QY 191 ALEDKIVEPHHHIGOTPAESDFOLLEIARLEMYGIRLHPAKDRGKTINLAVANTGIL 250  
 Db 183 MWEERITANYAEHRGRARDEAEWEYLKIAQDLEMYGVNYFAIRNKKGTLLGLVDALGLH 242  
 QY 251 VFQGGTKIN---AFNWKVKLSFKRKLRLKRPDANSAYODTLEFLMASRDFCKSFVK 307  
 Db 243 IYDPENRLPKISFPWNEIRNISYSDEFTIK-----PLDKKIDVFKFNSSKLRVNKLILQ 298  
 QY 308 ICVEHHAFF 316  
 Db 299 LCIGNHDLF 307  
 RESULT 14  
 US-08-179-738-7  
 ; Sequence 7, Application US/08179738  
 ; Patent No. 5578462  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Seizinger, Bernd R.  
 ; APPLICANT: Kley, Nikolai A.  
 ; APPLICANT: Bianchi, Albert B.  
 ; TITLE OF INVENTION: No. 5578462el NF2 Isoforms  
 ; NUMBER OF SEQUENCES: 26  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Reed & Robins  
 ; STREET: 635 Bryant Street  
 ; CITY: Palo Alto  
 ; STATE: California  
 ; COUNTRY: U.S.A  
 ; ZIP: 94301  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/179,738  
 ; FILING DATE: 10-JAN-1994  
 ; CLASSIFICATION: 530  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Robins, Roberta L.  
 ; REGISTRATION NUMBER: 33,208  
 ; REFERENCE/DOCKET NUMBER: 5998-0017  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 617-8999  
 ; TELEFAX: (415) 327-3231  
 ; INFORMATION FOR SEQ ID NO: 7:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 584 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; FRAGMENT TYPE: N-terminal  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: murine  
 ; US-08-179-738-7



```

; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: murine
US-08-628-145-7

Query Match      16.6%; Score 325; DB 1; Length 584;
Best Local Similarity 26.9%; Pred. NO. 2.8e-25;
Matches 83; Conservative 79; Mismatches 119; Indels 28; Gaps 9;

QY 23 STLERQKPPPTPSGKLVSISKIQLDDTQAEVFPORAPGKVLDDAVCNHNLNVEGDYFG 82
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12 SSLKR--KQPKFTTVRIVTM-----DAEMFNCMKWKCKDLFDLVCRTLGLRDTWFFG 63
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 83 LEPPDHKKITVWLDLLKPIV-KQIRRPKHVVVVKVFFPPD-HTOLQBELTRYLFALQV 140
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 64 LQY-TIKDTAVWLKMDKVLDDHVDVSKKEEPTVTHFLAKFPENAEELVQEIQTQLFFLQV 122
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 141 KODLAQGRLLTCNDTSAALLISHIVQSEIGDFDEAL-DREHLAKNKYIPQO-----D 190
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 123 KKEILDEKVCYCPPEASVLLASYAVQAKYGDYDPSVHKRGFLAQEELLPKRVINLYQWTP 182
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 191 ALEDKIVEFHNNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 183 MWEERTANYAEHRGRARDEAEYELKIAQDLEMYGVNYFTIRNKGTELLGLVDALGLH 242
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 251 VFOGFTKIN---AFNNAKVKLSFKRKRFLIKLRPDANSAYQDTLFLMASRDFCKSFWK 307
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 243 IYDPENRLTPKISFPWNEIRNISYSKKEFTIK----PLDKKIDVFKFDSSKLRVNLILQ 298
   :|:| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 308 ICVEHHAFF 316
   :|:| | | |
Db 299 LCIGNHDLF 307

RESULT 15
US-08-628-145-7
; Sequence 7, Application US/08628145
; Patent No. 5872214
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kley, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5872214el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,145
; FILING DATE: 04-APR-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,738
; FILING DATE: 10-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 584 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

Search completed: December 6, 2001, 08:55:09  
Job time: 291 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.  
OM protein - protein search, using sw model  
Run on: December 6, 2001, 08:52:34 ; Search time 44.17 Seconds  
(without alignments)  
644,992 Million cell updates/sec  
Title: US-09-555-342A-2\_COPY\_1\_374  
Perfect score: 1956  
Sequence: 1 MGEIEQRPTPGSLGAPENS.....GHKKVQFERKHSKIHSIRSL 374  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries  
Database : PIR.68:\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1949	99.6	1045	2	JC5795	CDEP protein - hum
2	802.5	41.0	1111	2	T23047	hypothetical prote
3	681.5	34.8	850	1	MMHUE4	erythrocyte membra
4	678.5	34.7	926	1	A41105	protein-tyrosine-p
5	666	34.0	858	2	A46613	protein 4.1, P4.1
6	661.5	33.8	1698	2	T13800	coracine gene prote
7	658	33.6	801	2	A37353	membrane protein 4
8	610	31.2	1026	2	T19631	hypothetical prote
9	527	26.9	568	2	T25859	hypothetical prote
10	506.5	25.9	913	1	A41109	protein-tyrosine-p
11	452.5	23.1	554	2	JU0188	membrane protein 4
12	423	21.6	639	2	T20772	hypothetical prote
13	423	21.6	4549	2	T20771	hypothetical prote
14	423	21.6	4667	2	T20774	hypothetical prote
15	368	18.8	1175	2	S51005	protein-tyrosine-p
16	368	18.8	1176	2	T58345	protein tyrosine p
17	362.5	18.5	577	1	A41289	moesin - human
18	361.5	18.5	577	1	S39804	moesin - pig
19	359	18.4	630	2	T47177	hypothetical prote
20	356	18.2	583	1	A46127	radixin - human
21	356	18.2	583	1	S39805	radixin - pig
22	356	18.2	583	1	A41129	radixin - mouse
23	356	18.2	1174	2	T38140	protein-tyrosine-p
24	351	17.9	581	2	T45889	ezrin - bovine
25	349	17.8	586	1	B41129	ezrin - mouse
26	348	17.8	586	1	A34400	ezrin [validated]
27	328	16.8	595	2	S33809	neurofibromin 2 -
28	327.5	16.7	654	2	T30957	hypothetical prote
29	327	16.7	591	2	I54368	merlin protein - m

30	327	16.7	596	2	I68664	merlin - mouse				
31	319	16.3	563	2	T29262	hypothetical prote				
32	313	16.0	564	2	T29263	hypothetical prote				
33	302	15.4	1189	1	JC2366	protein-tyrosine-p				
34	300	15.3	1187	1	JC4155	protein-tyrosine-p				
35	287	14.7	559	2	A45620	cyto villin homolog				
36	286	14.6	559	2	S49143	EG10 protein - tap				
37	274	14.0	2490	1	A54971	protein-tyrosine-p				
38	268	13.7	2450	2	S71625	protein tyrosine p				
39	265	13.5	2294	2	I67630	protein tyrosine p				
40	265	13.5	2466	2	I67629	protein tyrosine p				
41	205	10.5	90	2	A27056	erythrocyte membra				
42	155.5	7.9	652	2	T20549	hypothetical prote				
43	120	6.1	2491	2	A57036	talain - slime mold				
44	116	5.9	2541	2	S11661	talain - mouse				
45	106	5.4	444	2	T42674	hypothetical prote				

ALIGNMENTS

RESULT 1

JC5795

CDEP protein - human

C:Species: Homo sapiens (man)

C:Date: 24-Jan-1998 #sequence\_revision 13-Mar-1998 #text\_change 16-Jul-1999

C:Accession: JC5795

R:Koyano, Y.; Kawamoto, T.; Shen, M.; Yan, W.; Noshiro, M.; Fujii, K.; Kato, Y.

Biochem. Biophys. Res. Commun. 241, 369-375, 1997

A:Title: Molecular cloning and characterization of CDEP, a novel human protein containge factors.

A:Reference number: JC5795; MUID:98086358

A:Accession: JC5795

A:Molecule type: mRNA

A:Residues: 1-1045 <KOY>

A:Cross-references: DBJ:AB008430

C:Comment: This protein is involved in the adhesion, proliferation, and differentiation

C:Superfamily: pleckstrin repeat homology; protein 4.1 membrane-binding domain.homolo

F:1-374/Domain: ezrin-like #status predicted <EZR>

F:42-316/Domain: protein 4.1 membrane-binding domain homology <B41>

F:931-1027/Domain: pleckstrin repeat homology <PLK>

Query Match 99.6%; Score 1949; DB 2; Length 1045;

Best Local Similarity 99.7%; Pred. No. 2.1e-145;

Matches 373; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MGEIEQRPTPGSLGAPENSIGSTLERGOKPPPTPSGKLVSIKIQMLDDTQEAFAEVPQRA 60

Db 1 MGEIEQRPTPGSLGAPENSIGSTLERGOKPPPTPSGKLVSIKIQMLDDTQEAFAEVPQRA 60

Qy 61 PGKVLDDAVCNHNLNVEGDYFGLEFPDHRKKTIVWLDLLKPIVKQIRRPKHVVVVFYVKFF 120

Db 61 PGKVLDDAVCNHNLNVEGDYFGLEFPDHRKKTIVWLDLLKPIVKQIRRPKHVVVVFYVKFF 120

Qy 121 PPDHTQLOBELFRYLFAVQKODLAQGRLLTCNDTSAALISHIVQSEIGDFEALDREHL 180

Db 121 PPDHTQLOBELFRYLFAVQKODLAQGRLLTCNDTSAALISHIVQSEIGDFEALDREHL 180

Qy 181 ANKKYIPQDDALEDKIVEPHHNIHGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKI 240

Db 181 ANKKYIPQDDALEDKIVEPHHNIHGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKI 240

Qy 241 NLAVANTGLTVFQGTTKINAFNWKVKLSFKRKLRLKLRPDANSAYQDTLEFLMASRD 300

Db 241 NLAVANTGLTVFQGTTKINAFNWKVKLSFKRKLRLKLRPDANSAYQDTLEFLMASRD 300

Qy 301 FCKSEFWKICVEHAFRLFEPEPKPKPVLFSGSGSFRSGRTQKQVLDYVKEGGHKKVQ 360

Db 301 FCKSEFWKICVEHAFRLFEPEPKPKPVLFSGSGSFRSGRTQKQVLDYVKEGGHKKVQ 360

Qy 361 FERKHSKIHSIRSL 374

|||||

Db 361 FERKSHKHSIRSL 374

RESULT 2

T23047

hypotheetical protein H05G16.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000

C:Accession: T23047

R:White, S.

submitted to the EMBL Data Library, June 1997

A:Reference number: T19661

A:Accession: T23047

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1111 <W1>

A:Cross-references: EMBL:297190; PIDN:CAB10024.1; GSPDB:GN00028; CESP:H05G16.1

A:Experimental source: clone H05G16

C:Genetics:

A:Gene: CESP:H05G16.1

A:Map position: X

A:Introns: 97/3; 139/2; 171/3; 213/2; 294/3; 348/2; 373/2

Query Match 41.0%; Score 802.5; DB 2; Length 1111;

Best Local Similarity 44.2%; Pred. No. 4.8e-55;

Matches 165; Conservative 64; Mismatches 113; Indels 31; Gaps 6;

QY 22 ISVLEGRQKPPPPS-----GKLVSIKIQMLDDTQEAFF----- 55

DB 1 MSNIPRGVAGAPPPGMSNKKRGLVCIKVRLDDTVAVFHLGVICFFCVVSPHGYVFL 60

QY 56 VPOAPGKVLDDAVCNHLNVEGDYGLFEPDHPKHKTIVLWLLKPIVKQIRR-PKHVVVK 114

DB 61 LQKALGQTLLEDCVCHRLNLECDYGLSLFDINGNHCWLDREKTLIRQTNGSTDAKY 120

QY 115 FVYFFPPDHTQLQELTRYLFALQVKQDLAOGRLTCNDTSAAALLSHIVQSGIGDF--D 172

DB 121 FVYKFTPNPIDLEEYTRYLFMQIKRDLALGELCHSDNTASLLSAYLVQSGCGDFSSE 180

QY 173 EALDREHLAKNKYIQQD-ALEDKIVFEFHNNHIGQTPASDQQLLEIARLEMYGRLHP 231

DB 181 DYPDATYLSHTRVPVQTLFQKKVMDNHRNFTGMTFGESDLAMLEVARRCDFYGVKLA 240

QY 232 AKDREGTKINLAVANTGILVFOGFTKINAFNNAKVRLKSPKRRKRLIKLRPDANSAYQDT 291

DB 241 AKDIGNDAALSVMHLGIKVFRLQDLDTFSWARIKRLSKRKLKLVKLHPSYQYLKET 300

QY 292 LEFLMASRDFCKSFVKICVEHHAFFRLFEPPKPKPVLF-SRGSFSPSGRTQKQVLDY 350

DB 301 VEFSEFTRDECKNFKKVEHHAFFRCVQAEPEPKETREFFISKSSFRYHGRQKOLIDY 360

QY 351 VKEGGHKKVQFER 363

DB 361 VREHKKRREFPTR 373

RESULT 3

MMHUE4

erythrocyte membrane protein 4.1, parent splice form - human

N:Contains: erythrocyte membrane protein 4.1; lymphocyte membrane protein 4.1

C:Species: Homo sapiens (man)

C:Date: 31-Mar-1990 #sequence\_revision 08-Feb-1996 #text\_change 22-Jun-1999

C:Accession: A39810; A60244; A30207; A94143; A31260; B31260; G31260; S13152; A92559; A34377

R:Conboy, J.G.; Chan, J.Y.; Chasis, J.A.; Kan, Y.W.; Mohandas, N.

J. Biol. Chem. 266, 8273-8280, 1991

A:Title: Tissue- and development-specific alternative RNA splicing regulates expression of erythrocyte membrane protein 4.1

A:Reference number: A39810; MUID:91217063

A:Accession: A39810

A:Molecule type: mRNA

A:Residues: 1-850 <CON>

A:Cross-references: GB:M61733

R:Tang, T.K.; Leto, T.L.; Marchesi, V.T.; Benz Jr., E.J.

Adv. Exp. Med. Biol. 241, 81-95, 1988

A:Title: Expression of specific isoforms of protein 4.1 in erythroid and non-erythroid cells

A:Reference number: A60244; MUID:89132003

A:Accession: A60244

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-50, 'O', '52', 'L', '54-167', 'F', '169-227', '263-615', '656-787', 'K', '789-850' <TAN>

A:Experimental source: T-cell leukemia line MOLT-4

A>Note: lymphocyte membrane protein 4.1

R:Tang, T.K.; Leto, T.L.; Correas, I.; Alonso, M.A.; Marchesi, V.T.; Benz Jr., E.J.

Proc. Natl. Acad. Sci. U.S.A. 85, 3713-3717, 1988

A:Title: Selective expression of an erythroid-specific isoform of protein 4.1

A:Reference number: A94196; MUID:88234496

A:Accession: A30207

A:Molecule type: mRNA

A:Residues: 210-227, 263-615, 656-787, 'K', '789-850' <TA2>

A:Cross-references: GB:J03796; NID:g182072; PIDN:AAA35794.1; PID:g182074

A:Experimental source: T-cell leukemia line MOLT-4

A>Note: the authors translated the codon AAA for residue 539 as Gln

R:Conboy, J.; Kan, Y.W.; Shohet, S.B.; Mohandas, N.

Proc. Natl. Acad. Sci. U.S.A. 83, 9512-9516, 1986

A:Title: Molecular cloning of protein 4.1, a major structural element of the human erythrocyte membrane

A:Reference number: A94143; MUID:87092279

A:Accession: A94143

A:Molecule type: mRNA

A:Residues: 210-615, 635-758, 793-850 <CO2>

A:Cross-references: GB:M14993; NID:g182075; PIDN:AAA35795.1; PID:g182076

A:Experimental source: reticulocyte

R:Conboy, J.G.; Chan, J.; Mohandas, N.; Kan, Y.W.

Proc. Natl. Acad. Sci. U.S.A. 85, 9062-9065, 1988

A:Title: Multiple protein 4.1 isoforms produced by alternative splicing in human erythrocytes

A:Reference number: A94215; MUID:89057876

A:Accession: A31260

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 615, 656-700 <CO3>

A:Experimental source: reticulocyte

A:Accession: B31260

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 701-714, 738-837, 'L', '839-848', 'E' <CO4>

A:Experimental source: reticulocyte

A:Accession: C31260

A:Status: nucleic acid sequence not shown; not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 701-758, 793-837, 'L', '839-848', 'E' <CO5>

A:Experimental source: reticulocyte

R:Horne, W.C.; Prinz, W.C.; Tang, E.K.Y.

Biochim. Biophys. Acta 1055, 87-92, 1990

A:Title: Identification of two CAMP-dependent phosphorylation sites on erythrocyte protein 4.1

A:Reference number: S13152; MUID:91027920

A:Accession: S13152

A:Molecule type: protein

A:Residues: 534-541; 693-701; 758, 793-794 <HOR>

R:Correas, I.; Speicher, D.W.; Marchesi, V.T.

J. Biol. Chem. 261, 13362-13366, 1986

A:Title: Structure of the spectrin-actin binding site of erythrocyte protein 4.1

A:Reference number: A92559; MUID:87008553

A:Accession: A92559

A:Molecule type: protein

A:Residues: 615, 635-700 <COR>

A>Note: this 67-residue peptide forms a ternary complex with spectrin and actin; anti-peptide antibodies inhibit the complex

R:Inaba, M.; Maeda, Y.

J. Biol. Chem. 264, 18149-18155, 1989

A:Title: O-N-acetyl-D-glucosamine moiety on discrete peptide of multiple protein 4.1

A:Reference number: A34377; MUID:90036892

A:Accession: A34377

A:Molecule type: protein

A:Residues: 772-791 <INA>

A>Note: this peptide region appears to have several partially glycosylated Ser and Thr residues

C:Comment: In mammalian erythrocytes, protein 4.1 stabilizes the spectrin-actin network band 3 and glycophorin.

C:Comment: Four domains are proposed based on the properties of peptides released after 3, glycoprotein, and phospholipids. This domain is hydrophobic, and has two potential high in proline.  
 C:Comment: Following the highly-charged domain that binds spectrin and actin (residues 6-11), the boundary between the membrane-binding and hydrophilic domains is uncertain.  
 C:Genetics:

A:Gene: GDB:EPB41

A:Cross-references: GDB:119865; OMIM:130500

A:Map position: 9q31-q931

C:Superfamily: membrane protein 4.1; protein 4.1 membrane-binding domain homology

C:Keywords: alternative initiators; alternative splicing; cytoskeleton; erythrocyte; MEM

F:1-227,263-615,656-850/Product: lymphocyte membrane protein 4.1 #status predicted <MEM>

F:210-615,635-758,793-850/Product: erythrocyte membrane protein 4.1 #status predicted <MEM>

F:210-487/Domain: membrane-bound #status predicted <DM2>

F:212-487/Domain: protein 4.1 membrane-binding domain homology <B41>

F:488-614/Domain: hydrophilic <DM2>

F:615-700/Domain: spectrin/actin binding #status predicted <DM3>

F:701-850/Domain: carboxyl-terminal <DM4>

F:540,695/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status ex

Query Match 34.8%; Score 681.5; DB 1; Length 850;  
 Best Local Similarity 42.0%; Pred. No. 1.1e-45;  
 Matches 152; Conservative 58; Mismatches 143; Indels 9; Gaps 7;

QY 16 APENGISTLERGQPPPTPPSKGLSVIKIOMLDDTQAEFVQAPGKVLDDAVCNHNL 75

Db 188 SPQSAETELKASQK--PIRKHRNHCKVSLDDTVYECWEKHAQGDQLLRVCHEHNL 245

QY 76 VEGDYFGLFEPDHKKITVWDLKLPVQIRK--PKHVVVVFVVKFPPDPHTQOELTRY 134

Db 246 LDEDFGLAIWNAISKVLSAKELKQVGVNFT--FNVKFPDPAQLTEDITRY 303

QY 135 LFALQVQDLAQRGRLTCDNTSAALISHIVQSEIGDFDEALDR--EHLAKNKYIPQO--DAL 192

Db 304 YLCQLRQDVRAGRPCSFATLALLGSYTIQSELGDDYDPELHGVDYVDFKLAPNOTKEL 363

QY 193 EDKIVFHHHIGQTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILV 252

Db 364 EEKVELHKSYSRMTPAQADLEFLENAKKLSMYGVDLHKAKDLEGVDITLGVCSGLLYV 423

QY 253 OGFTKINAFNAKVKRLSPKRRFLIKLRPDANSAYODTLEFLMASRDFCKFWKICV 312

Db 424 KDKLINRPPVWKVUKISYRFFIKIRPGEQYETIGFKLPSYRAAKLWKVCV 483

QY 313 HAFRLFEPPKPKPVLSFRSGSSFRSGRTQKQVLDYVKEGHHKVKQPERKHSK 372

Db 484 HFFRL-TSTDIPKSKFLALGSKFRYSQRTQAQTRQASALIDRPAHPERTASK-RASR 541

QY 373 SL 374

Db 542 SL 543

RESULT 4

A:1105

protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 4 - human

N:Alternate names: PTPase MEG

C:Species: Homo sapiens (man)

C:Date: 20-Mar-1992 #sequence\_revision 02-May-1994 #text\_change 11-Jun-1999

C:Accession: A41105

R:Gu, M.; York, J.D.; Warshawsky, I.; Majerus, P.W.

Proc. Natl. Acad. Sci. U.S.A. 88, 5867-5871, 1991

A:Title: Identification, cloning, and expression of a cytosolic megakaryocyte protein-ty

A:Reference number: A41105; MUID:91288564

A:Accession: A41105

A:Molecule type: mRNA

A:Residues: 1-926 <GUA>

A:Cross-references: GB:M68941; NID:g190747; PIDN:AAA36530.1; PID:g190748

A:Experimental source: megakaryocytes, cell line MEG-10

C:Genetics:

A:Gene: GDB:PTN4

A:Cross-references: GDB:131387; OMIM:176878

A:Map position: 9q31-q931

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology  
 C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase  
 F:31-308/Domain: protein 4.1 membrane-binding domain homology <B41>

F:523-597/Domain: GLGF domain homology <GLG>

F:679-900/Domain: protein-tyrosine-phosphatase homology <PTP>

F:852/Active site: Cys (phosphocysteine intermediate) #status predicted

F:858/Binding site: substrate phosphate (Arg) #status predicted

Query Match 34.7%; Score 678.5; DB 1; Length 926;

Best Local Similarity 41.1%; Pred. No. 2.2e-45;

Matches 146; Conservative 62; Mismatches 124; Indels 23; Gaps 6;

QY 35 PSGLKLSIK-----IQMLDDTQAEFVQAPGKVLDDAVCNHNLVEG 78

Db 8 PAGRTYVNRASLARDRQHTVEVCNILLDNTVQAEKVNKHQDQVLLDVFVKHLDTQ 67

QY 79 DYFGLFEPDHKKITV-WDLKLPVQIRKPKHVVVVFVVKFPPDPHTQOELTRYLFA 137

Db 68 DYFGLQADDDSTNPRWLPNPKPIRKQKRGSPYSLNFRVKFVSDPNKLQEEYTRYQVF 127

QY 138 LQVKQDLAQRGRLTCDNTSAALISHIVQSEIGDFDEALDR--EHLAKNKYIPQO--DALEK 195

Db 128 LQVKQDLGRLPCPSNTAALLASFAVQSELGDDYDSENLGSLYSYFIPNQPDPEKE 187

QY 196 IVEFHHHIGQTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFOG 255

Db 188 IAKHQHIGLSPAAEFNLTARTLEGVFEVHFYARDQSNNEIMIGVMSGGILYKVR 247

QY 256 TKINAFNAKVKRLSPKRRFLIKLRPDANSAYODTLEFLMASRDFCKFWKICV 315

Db 248 VRMTTPWLKIVKISPKCKOFFQLKELHRESRETLGFMVNYRACKNLKACV 307

QY 316 ERLFEPPKPKPVLS--FSRSGSSFRSGRTQKQVLDYVKEGHHKVKQPERKHSK 367

Db 308 FRL-DRPLPPQKNFFAHYFTLGSKFRYCGRTVEQVQYQYKQKANKDRVFARSPSK 361

RESULT 5

A46613

protein 4.1, P4.1 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 25-Aug-1995

C:Accession: A46613

J:Huano, J.P.; Tang, C.J.; Kou, G.H.; Marchesi, V.T.; Benz Jr., E.J.; Tang, T.K.

J. Biol. Chem. 268, 3758-3766, 1993

A:Title: Genomic structure of the locus encoding protein 4.1. Structural basis for co

A:Reference number: A46613; MUID:93155238

A:Accession: A46613

A:Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-858 <HUA>

A:Experimental source: embryo

A:Note: sequence extracted from NCBI backbone (NCBI:124466, NCBI:124467)

C:Superfamily: membrane protein 4.1; protein 4.1 membrane-binding domain homology

F:213-488/Domain: protein 4.1 membrane-binding domain homology <B41>

Query Match 34.0%; Score 666; DB 2; Length 858;

Best Local Similarity 40.9%; Pred. No. 1.9e-44;

Matches 158; Conservative 53; Mismatches 155; Indels 20; Gaps 9;

QY 5 EORPTPGSRL---GAPENSGISTLE---RGOKP-----PTPSGLKLSIKIOMLDDTQ 51

Db 163 EHREDPDSETKEGEGIECSGTEVKEDPSRAEREPASQKPVRRHNRHCKVSLDDTV 222

QY 52 EAFVQAPGKVLDDAVCNHNLVEGDFGLFEPDHKKITVWDLKLPVQIRK--PKH 110

Db 223 YECVVEKHANLQDLKRVCEHLNLEEDYFGLALWDSATSKTLDLSAKETKKQVGPWN 282

QY 111 VVVKFVVKFPPDPHTQOELTRYLFAQVKQDLAQRGRLTCDNTSAALLISHIVQSEIGD 170

Db 283 FT--FNVKFPYPPDAQLTEDITRYICLQLQRQDIVAGRLPCSFATLALLGSIYTIQSELGD 340  
QY 171 FDEAL-DREHLAKNKVIPOQ-DALEDKIVEFHNNHIGQTPAESDFOLLEIARLEMYGIR 228  
Db 341 YDELFHGMVDVDFKLPAGNQTKELEKVMELHKSYSMTMPAQADLEFLENAKKLSMYGVD 400  
QY 229 LHPAKDREGTKINLAVANTGILVFOGFTKINAFNNAKVRKLSFKRFLIKLRPDANSAY 288  
Db 401 LHRKADLEGVDITLGVCSGLLYVKDLINRPPWPKVLKISSKRSFFIKIRPGEQEHY 460  
QY 289 QDTLEFLMASRDFCKSFWKTCVHHAFRLFEFPKPKPVLFPSRGSSFRFSQRTQKVL 348  
Db 461 ESTIGFKLPSYRAAKLWKVCEVHHFTFRL-TSTDTPKSKFLALGSKFRYSRGTQATR 519  
QY 349 DYVKEGCHKVQPERKHSKHSIRSL 374  
Db 520 QASALIDRPAHPERTASK-RASRSL 544  
RESULT 6  
T13800  
coracle gene protein - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: T13800  
R:Fehon, R.G.; Dawson, I.A.; Artavanis-Tsakonas, S.  
Development 120, 545-557, 1994  
A:Title: A Drosophila homologue of membrane-skeleton protein 4.1 is associated with septin  
A:Reference number: Z17763; MUID:94215495  
A:Accession: T13800  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1698 <FEH>  
A:Cross-references: EMBL:L27467; NID:g440293; PID:g440294; PIDN:AA859187.1  
C:Genetics: coracle  
A:Gene: coracle  
A:Cross-references: FlyBase:FBgn0010434  
A:Map position: 2  
C:Function:  
A:Description: may play a role in cell-cell interactions  
C:Keywords: alternative splicing  
Query Match 33.8%; Score 661.5; DB 2; Length 1698;  
Best Local Similarity 40.8%; Pred. No. 1.1e-43;  
Matches 142; Conservative 64; Mismatches 125; Indels 17; Gaps 7;  
QY 5 EQRPFGSLGAPENSIGTLERGQPPPTPS--GKLVSIKIQMLDDTQEAFFVPPORAPG 62  
Db 4 EIKPSAPAEPTPKS-----KPKSSSSHGKPALARVTLTDLGSLDVSIDRAKIG 54  
QY 63 KVLDDAVCNHLNLVDEGDFGLPEPDHKKITVWLDLLKPIVKQIRRPKHVVVKKVFFPP 122  
Db 55 RDVINSICAGLNLEIKDYGLIYETPDPTWLDLEKPVSKFRTDTWPLT-FAVAFYPP 113  
QY 123 DHTQLQELTRYLFALQVQDLAQGRLTCDNTSAALLSHIVQSEIGDFD--BALDREHL 180  
Db 114 EPSQLKEDITRYHLCQLQVRNILEGRLPCTFTVTHALLGSLVQSEMGDYDAEMPTRAVL 173  
QY 181 AKNKYIPOQDA-LEDKIVEFHNNHIGQTPAESDFOLLEIARLEMYGIRLHPAKDREGTK 239  
Db 174 KDFKIAPNQTAELEDDVMDLHKTHKQSPAEALHYLENAKKLAMYGVDLHPAKDSEGV 233  
QY 240 INLAVANTGILVFOGFTKINAFNNAKVRKLSFKRFLIKLRPDANSAYQDTLEFLMASR 299  
Db 234 IMGVCSAGLLVTRDKLRINRFAWPKILKISYKRRHHFYIKIRPGEFEQYESTIGFKLANH 293  
QY 300 DFCKSFWKICVHHAFRLFEFPKPKPVLFPS-RGSSFRFSQRTQK 346  
Db 294 RAAKLWKSVEHHTFRLM-TPEPYSKSMFPVFGSTYRKGRTOAE 340  
RESULT 7

A37353  
membrane protein 4.1 - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 19-Mar-1993 #sequence\_revision 19-Mar-1993 #text\_change 13-Aug-1999  
C:Accession: A37353; A29901  
R:Spencer, M.; Giebelhaus, D.H.; Kelly, G.M.; Bicknell, J.; Florio, S.K.; Milam, A.H.  
Dev. Biol. 139, 279-291, 1990  
A:Title: Membrane skeleton protein 4.1 in developing Xenopus: expression in postmitotic  
A:Reference number: A37353; MUID:90249600  
A:Accession: A37353  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-801 <SPE>  
A:Cross-references: GB:M20621; NID:g214090; PIDN:AAA49695.1; PID:g214091  
R:Giebelhaus, D.H.; Eib, D.W.; Moon, R.T.  
Cell 53, 601-615, 1988  
A:Title: Antisense RNA inhibits expression of membrane skeleton protein 4.1 during em  
A:Reference number: A29901; MUID:88223353  
A:Accession: A29901  
A:Molecule type: mRNA  
A:Residues: 1-550 <GIE>  
C:Superfamily: membrane protein 4.1; protein 4.1 membrane-binding domain homology  
C:Keywords: cytoskeleton; membrane protein  
F:195-470/Domain: protein 4.1 membrane-binding domain homology <B41>  
Query Match 33.6%; Score 658; DB 2; Length 801;  
Best Local Similarity 41.5%; Pred. No. 7.4e-44;  
Matches 139; Conservative 60; Mismatches 116; Indels 20; Gaps 6;  
QY 28 GQKPPPTPSGKL-----VSIIQMLDDTQEAFFVPPORAPGKVLDDAVCNHLNL 75  
Db 169 GREPIKKPEGESKASHKVVRRSPNMRCKVTLLDDTVVCECDLEKHAQGDIFKKVCSHLNI 228  
QY 76 VEGDYEGLEFPDHHKKTIVWLDLLKPIVKQIRRPKHVVVAFV--VKFPPDHTOLOEELTR 133  
Db 229 VEEDYEGLAIWESPTCKVWLDPLDKIRKQVHGPP---CEFTSNVKFYPPDPAQLSSEDIR 285  
QY 134 YLFALQVQDLAQGRLTCDNTSAALLSHIVQSEIGDFDEALDR-EHLAKNKYIPOQDA- 191  
Db 286 YILCLQRKIDFSGRLPCSFATLALLGSIYVQSEVGEEDLHGVDYVSEFKLSPNQTKD 345  
QY 192 LEDKIVEFHNNHIGQTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILV 251  
Db 346 LEEKVGEELHKSYSMTMPAQADLEFLENAKKLTMYGVDIHOAKDLEGVDIKLGVCSGLMV 405  
QY 252 FQGFTKINAFNNAKVRKLSFKRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVE 311  
Db 406 FKNLNRINRPPWPKVLKISYKRSFFIKIRPGEQYESTIGFKLPSYKAARKLWKVCVE 465  
QY 312 HHAFRLFEFPKPKPVLFPSRGSSFRFSQRTQK 346  
Db 466 HHTFFRL-TSTESIPKHFSLSLGSTRYSRGTQAK 499  
RESULT 8  
T19631  
hypothetical protein C48D5.2a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Jan-2000  
C:Accession: T19631; T20040  
R:Thomas, K.  
submitted to the EMBL Data Library, February 1995  
A:Reference number: Z19154  
A:Accession: T19631  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1026 <WIL>  
A:Cross-references: EMBL:Z48241; PIDN:CAA88287.1; GSPDB:GN00021; CESP:C48D5.2a  
R:Lightning, J.  
submitted to the EMBL Data Library, August 1994  
A:Reference number: Z19215

A:Accession: T20040  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1026 <W12>  
A:Cross-references: EMBL:Z36237; PIDN:CAAB5272.1; GSPDB:GN00021; CESP:C48D5.2a  
A:Experimental source: clone C48D5  
C:Genetics:  
A:Gene: CESP:C48D5.2a  
A:Map position: 3  
A:Introns: 46/3; 85/3; 159/1; 200/2; 282/3; 358/2; 396/2; 447/2; 481/3; 543/1; 605/3; 70/3  
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology; P

Query Match 31.2%; Score 610; DB 2; Length 1026;  
Best Local Similarity 39.1%; Pred. No. 6.3e-40;  
Matches 147; Conservative 60; Mismatches 145; Indels 24; Gaps 10;

QY 13 RLCAPEISGISTLERQKPPPTSGKLVSIKIQLMDTQEAPEVQAPGKVLVLDVACNH 72  
DB 2 RLGSNDYQVQTEAIGQTPVPPNQIRCTVTFDSTSYHFEIEKNSLGIVLLEKVPNY 61  
QY 73 LNLVEGDYFGLFE-----PDHKKITVWLDLKPVKO-IRPKHVVVKKFVVKFPPDH 124  
DB 62 LEIEKDYEGLVFIADVNSSAQKK---NLDPSKLNLRKQWICPPYHLF--FRVKFVVRDP 116  
QY 125 TQLOEELRYLALQVQDLAQLTCNDTSAALLISHIVQSEIGDFDE--ALDREHLA 181  
DB 117 NLRDFTFRFQYQVQRNLEGRPCNGLSALLASYVYVQAEVGFEEKTHGMSRTCLC 176  
QY 182 -KNKIYPOQDALEKIVFHHNHIGQTPAESDFOLLEIARRHLEMYGIRLHPAKDREGTKI 240  
DB 177 YKIQFATLPDSDRVAELHQLHIGQTPDVAQONFLDHARRLEMYGMDVYDGVNHLPI 236  
QY 241 NLAVANTGILVFGFTKINAFNMAKVRKLSFRKRLIKLRPDANSAYQOTLEFLMASRD 300  
DB 237 EIGVGAVGKIVHEGKIMNEYAVWIRKLSFKKKQVQ-VLANEDGVSETIMIFNMSAK 295  
QY 301 FCKSFVKICVEHHAFLRPEEPKPKVLFSGSSFRSGRTQKQVL---DYVKEGGHK 357  
DB 296 ICKLLMKCKIEQHTFRLTPPK-TPQRKVFNGSKFRYSGRTEYQTLLENEHRKSAGHR 354  
QY 358 KVOFERKSHKSHSIR 373  
DB 355 --NFRSLSKSFLRS 368

RESULT 9  
T25859  
Hypothetical protein T04C9.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T25859  
R:Favella, A.  
Submitted to the EMBL Data Library, December 1996  
A:Description: The sequence of C. elegans cosmid T04C9.  
A:Reference number: Z20101  
A:Accession: T25859  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-568 <FAV>  
A:Cross-references: EMBL:U80955; PIDN:AAB38104.1; GSPDB:GN00021; CESP:T04C9.6  
A:Experimental source: strain Bristol N2; clone T04C9  
C:Genetics:  
A:Gene: CESP:T04C9.6  
A:Map position: 3  
A:Introns: 52/3; 87/3; 129/2; 162/1; 197/2; 280/3; 393/3; 538/3

Query Match 26.9%; Score 527; DB 2; Length 568;  
Best Local Similarity 35.9%; Pred. No. 9.7e-34;  
Matches 125; Conservative 64; Mismatches 139; Indels 20; Gaps 8;

QY 38 KLVSIKIQMLDQEAFAFVQAPGKVLVLDVACNHLNIVEGDYFGLFEPDHKKITVWLDL 97

Db 33 KFVQCKVLLDGAHLNIVVPRNAGSELVEEYFSLDLERDYFGLOQYTDFFNVQHWLDP 92  
QY 98 LKPIVAKIIRPRKHVVVVKVVFPPD-HTQLQBELRYLALQVQKDLAQLRITCNDTSA 156  
DB 93 TKVAKQVAIGPPPTLRFVRKFPFTSPSSNLKEELRYQFFLQIKQDISSGRQCQPHLA 152  
QY 157 ALLISHIVQSEIGDFDEALDRE-HLAKNKYIPOQD-ALEDKIVFHHNHIGQTPAESDFQ 214  
DB 153 IEIAAFALSGELGYNPELHTALFISEFRPHPEQDEKMEVEILE-----SGQTPAQALN 207  
QY 215 LLEIARRLEMYGIRLHPAKDREGTKINLAVANTGILVFOGFTKINAFNMAKVRKLSFKRK 274  
DB 208 YLNKARWIEYGVDMHIVEGKDGNTYRLGLTPQGLMVLVDFGPKIGLFEKLGKLDKPKK 267  
QY 275 RELIKLRPDANSAYQ-----DTLEFLMASRDFCKSEKFWKICVEHHAFLRPEEPKPKPV 329  
DB 268 KITLVVEEDADQSNNGIOQLHTEVHLTSEKAAKHEWKCAIEQAFAFLKLSRPIQANRKI 327  
QY 330 LFSR-CSSFRFSRGTOKQVLDYVKEG---HKKVQFERKSHKSHSIR 372  
DB 328 QFFRLGSTFKYGRTEYETIH--KEGARLSRRQSCSFERRPSQRYGPR 373

RESULT 10  
A1109  
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 3 - human  
N:Alternate names: PTPH1  
C:Species: Homo sapiens (man)  
C:Date: 27-Mar-1992 #sequence\_revision 02-May-1994 #text\_change 11-Jun-1999  
C:Accession: A1109; 155698  
R:Yang, Q.; Tonks, N.K.  
Proc. Natl. Acad. Sci. U.S.A. 88, 5949-5953, 1991  
A:Title: Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with  
A:Reference number: A1109; MUID:91296738  
A:Accession: A1109  
A:Molecule type: mRNA  
A:Residues: 1-913 <IAN>  
A:Cross-references: GB:I64572; NID:9179912; PIDN:AAA35647.1; PID:9179913  
R:Ikuta, S.; Itoh, F.; Hinoda, Y.; Toyota, M.; Makiguchi, Y.; Imai, K.; Yachi, A.  
J. Gastroenterol. 29, 727-732, 1994  
A:Title: Expression of cytoskeletal-associated protein tyrosine phosphatase PTPH1 mRNA  
A:Reference number: 155698; MUID:95179278  
A:Accession: 155698  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 899-913 <RES>  
A:Cross-references: GB:S76309; NID:913165; PIDN:AAB33583.1; PID:913166  
C:Genetics:  
A:Gene: GDB:PTPN3  
A:Cross-references: GDB:131386; OMIM:176877  
A:Map position: 9q31-9q31  
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology  
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase  
F:31-308/Domain: protein 4.1 membrane-binding domain homology <B41>  
F:516-590/Domain: GLGF domain homology <GLG>  
F:670-890/Domain: protein-tyrosine-phosphatase homology <PTP>  
F:842/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:848/Binding site: substrate phosphate (Arg) #status predicted

Query Match 25.9%; Score 506.5; DB 1; Length 913;  
Best Local Similarity 35.0%; Pred. No. 7.6e-32;  
Matches 119; Conservative 68; Mismatches 134; Indels 19; Gaps 9;

QY 27 RGOKPPPTSGKLVSIKIQLMDTQEAFAFVQAPGKVLVLDVACNHLNIVEGDYFGLFEP 86  
DB 16 RSELPKTRSEVICSIHFLDGVGVTFKVTQDTQGVLDVMVHNLGVTEKEFYFGLQHD 75  
QY 87 DHKKITV-WLDLKLKPIVKQIRPRKHVVVVKVVFPPDHTQLQBELRYLALQVQKDLA 145  
DB 76 DSDVDSFRLWEASKPIRKQLKGGFPCTLHFRVRFPPDPNTLQEQTRHLYFLQLKWDIC 135

QY 146 QGRLTNDTSAALLSHIVQSEIGDDEALDRE-HLAKNKYIPQDALED---KIVEFHH 201  
Db 136 EGRLTCPNLNAVYASVAVQSHFGDYNSSHTHPGYLSDSHFIPDQN--EDFLTKVESLHE 193  
QY 202 NHGQTPAESDFOLLETAARLEMYGIRLHPAKDREGTKINLAVANTGILVFGQFTKINAF 261  
Db 194 QHSGLKQSEAESYINARTLDYGVHELHSGRDLNLDLMIGTASGAVAYRYKICTSFY 253  
QY 262 NNAKVRKLSFKRRFKIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHAFFRLFE 321  
Db 254 PWNILKISFKRKKFIHQKQKQAESREHVAENMLNRYCKNLKSCVEHHTFFQK-KK 312  
QY 322 PKPKPVL---PSRGSFRSFGRTQKQVLD-VYKE--GG 355  
Db 313 LLPQKNVLQSYWTMGSR-----NTRKSVNNQCKKVIIG 347  
RESULT 11  
J00188  
membrane protein 4.1 homolog - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Feb-1994 #sequence\_revision 20-Aug-1994 #text\_change 05-Nov-1999  
C:Accession: I55505; J00188  
R:Takeuchi, K.; Kawashima, A.; Nagafuchi, A.; Tsukita, S.  
J. Cell Sci. 107, 1921-1928, 1994  
A:Title: Structural diversity of band4.1 superfamily members.  
A:Reference number: I55505; MUID:95074267  
A:Accession: I55505  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-534 <RES>  
A:Cross-references: GB:D28818; NID:9466547; PIDN:BAA05978.1; PID:9466548  
C:Superfamily: protein 4.1 membrane-binding domain homology  
F:13-295/Domain: protein 4.1 membrane-binding domain homology <B41>  
Query Match 23.1%; Score 452.5; DB 2; Length 554;  
Best Local Similarity 33.1%; Pred. No. 6.9e-28;  
Matches 105; Conservative 65; Mismatches 132; Indels 15; Gaps 7;  
QY 43 KIOMLDDTQAEFVPQOR-----APGKVLDDAVCNHNLNVEGDYFGLFEPDHHKTIWLDL 97  
Db 14 EVLLDESKLTLTQOQGIKSTKGSVLDHVRHNLNVEIDYFGLRYCDRSHQTYWLDP 73  
QY 98 LKPIVKQ---IRPKHVVVKKVFPDPDHTLOELTRYLQALQVQDLQACRLTCNDT 154  
Db 74 AKTLAEKELINTGPPYTLVFGIKFAEDPCKLEIITRYQFLQVQDALQGRLPVYN 133  
QY 155 SAALLSHIVQSIGDFDEALDRE-HLAKNKYIP-QQDALEDKIVFHHNHIGQTPAESD 212  
Db 134 IAAQMGAYAIQABLDHDPYKHTAGVYSEVRFVDPQKEELEAIERIHKTLMGQAPSEAE 193  
QY 213 FOLLETAARLEMYGIRLHPAKDREGTKINLAVANTGILVFGQFTKINAFNNAKVRKLSFK 272  
Db 194 LNLRTAKSLEMYGVDLHPYGENKSEYFLGTPSGVVYVKKQVKGKFWPRITKVFHK 253  
QY 273 RKFLIK-LRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHAFFRLFEKPKPVLV 331  
Db 254 ETQFELRVLGKDCN---EFSFFEASKTACKHLWKSCVEHHTFFRMPDTESNLSRKLS 310  
QY 332 SRGS-SFRSFGRTQKV 347  
Db 311 KFGSISYKHYRTALQM 327  
RESULT 12  
T20772  
hypothetical protein ZK270.2b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T20772; T27815  
R:Kershaw, J.  
submitted to the EMBL Data Library, November 1996

A:Reference number: Z19322  
A:Accession: T20772  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-639 <WIL>  
A:Cross-references: EMBL:281499; PIDN:CAB04091.1; GSPDB:GN00019; CESP:ZK270.2b  
A:Experimental source: clone F11C3  
R:Kershaw, J.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z20423  
A:Accession: T27815  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-639 <W12>  
A:Cross-references: EMBL:282089; PIDN:CAB05004.1; GSPDB:GN00019; CESP:ZK270.2b  
A:Experimental source: clone ZK270  
C:Genetics:  
A:Gene: CESP:ZK270.2b  
A:Map position: 1  
A:Introns: 61/3; 112/1; 244/3; 328/2; 384/1; 433/3; 467/2; 559/3; 610/3  
Query Match 21.6%; Score 423; DB 2; Length 639;  
Best Local Similarity 33.8%; Pred. No. 1.8e-25;  
Matches 113; Conservative 58; Mismatches 145; Indels 18; Gaps 10;  
QY 28 GOKPPPTPS-----GKLVSIKIQMLDDTQAEFVPQAPGKVLDDAVCNHNLNVEGDYFG 82  
Db 27 GDQKHTTQQPRDSKWLAKVLLPDGVQKEFEVKNSEGEALFROVTRDLSIEEYFYS 86  
QY 83 LEFPDHHKIT-VMLDILLKPIVKQIR-RPRHVVVKKVFPDPDHTLOELTRYLQALQV 140  
Db 87 LCFYDKDEGTRHLYNDKNIAQIKGLPWE--FSPEVKFVPTTPTTIVDDHARYVYVFLQL 144  
QY 141 KQDLAAGRLTCNDTSAALLSHIVQSEIGDFD-EALD--REHLAKNKYIPQDALED--- 194  
Db 145 RDDLTLGRLPATADTSHLGSFVAQIEFGDAPAEMTDAYEQFIVASKLPSPAQNPETYK 204  
QY 195 KIVEFHNNHIGQTPAESDFOLLETAARLEMYGIRLHPA-KDREGTKINLAVANTGILVQ 253  
Db 205 KIVDLHREMRGQTPSEAEHQFDHCKHLALYGIHLFKALSDKDKKPVGVGIGAGINIQ 264  
QY 254 GFTKINAFNNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHH 313  
Db 265 DEQKTHSFQWNLKIGYRRTVFSIKLQAGTVEKNEKTLVFKLPNHVAAKRTWKCAVEHH 324  
QY 314 AFRLFEKPKPVPVLSRGS--FRFSGRTOQK 346  
Db 325 TFFRLT-QPEDKTHKSFNFQSGRFRYQGRTOFQ 357  
RESULT 13  
T20771  
hypothetical protein ZK270.2a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T20771; T27814  
R:Kershaw, J.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19322  
A:Accession: T20771  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4549 <WIL>  
A:Cross-references: EMBL:281499; PIDN:CAB04090.1; GSPDB:GN00019; CESP:ZK270.2a  
A:Experimental source: clone F11C3  
R:Kershaw, J.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z20423  
A:Accession: T27814  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4549 <W12>





Db 186 QKVALHQQYRGLTAPAEMLYMOEVERMDGYGEESYPKADSQGSDISIGACLDGIFVKH 245  
Qy 252 FQGETKINAFNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSEFWKICVE 311  
Db 246 KNGRPPV-VFRWHDIANNSHNSFFALEL--ANK--EETIQOTEDMETAKYVWRLCVA 299  
Qy 312 HHAFRLFEFPKPKPKPVLS--RGSSFRFS 340  
Db 300 RHKEYRLNQC�LQTAATLNSVRGRSSRMS 330

Search completed: December 6, 2001, 08:52:37  
Job time: 369 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:53:07 ; Search time 28.69 Seconds  
(without alignments)  
477.959 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_1\_374

Perfect score: 1956

Sequence: 1 MGEIEQRPTGSLGAPENS.....GHKKVQFERKXKIHISRLS 374

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	681.5	34.8	864	1 41_HUMAN	P11171 homo sapien
2	678.5	34.7	926	1 PTV4_HUMAN	P29074 homo sapien
3	669.5	34.2	881	1 E4L1_HUMAN	Q9H490 homo sapien
4	666	34.0	858	1 41_MOUSE	P48193 mus musculus
5	658	33.6	801	1 41_XENLA	P11434 xenopus lae
6	610	31.2	1026	1 PTP1_CAEEL	P28191 caenorhabdi
7	587.5	30.0	732	1 YF48_HUMAN	Q9Hcm4 homo sapien
8	506.5	25.9	913	1 PTV3_HUMAN	P26045 homo sapien
9	496	25.4	619	1 NBL4_BRARE	O57457 brachydanio
10	480.5	24.6	598	1 NBL4_HUMAN	Q9Hcs5 homo sapien
11	452.5	23.1	554	1 NBL4_MOUSE	P52963 mus musculus
12	368	18.8	1175	1 PTVL_RAT	Q62728 rattus norv
13	368	18.8	1176	1 PTVL_MOUSE	Q62136 mus musculus
14	363.5	18.6	576	1 MOES_MOUSE	P26041 mus musculus
15	362.5	18.5	576	1 MOES_HUMAN	P26038 homo sapien
16	361.5	18.5	576	1 MOES_PIG	P26042 sus scrofa
17	356	18.2	583	1 RADI_HUMAN	P35241 homo sapien
18	356	18.2	583	1 RADI_MOUSE	P26043 mus musculus
19	356	18.2	583	1 RADI_PIG	P26044 sus scrofa
20	356	18.2	1174	1 PTVL_HUMAN	Q16825 homo sapien
21	351	17.9	580	1 EZRI_BOVIN	P31976 bos taurus
22	349	17.8	585	1 EZRI_MOUSE	P26040 mus musculus
23	348	17.8	585	1 EZRI_HUMAN	P15311 homo sapien
24	328	16.8	595	1 MERL_HUMAN	P35240 homo sapien
25	327	16.7	596	1 MERL_MOUSE	P46662 mus musculus
26	306.5	15.7	572	1 MOES_LYTV	P52962 lytechinus
27	302	15.4	1189	1 PTVN_MOUSE	Q62130 mus musculus
28	300	15.3	578	1 MOEH_DROME	P46150 drosophila
29	300	15.3	1187	1 PTVN_HUMAN	Q15678 homo sapien
30	276.5	14.1	1051	1 YC94_HUMAN	Q9P292 homo sapien
31	265	13.5	2485	1 PTVN_HUMAN	Q12523 homo sapien
32	251	12.8	1062	1 YAI3_HUMAN	Q9Y216 homo sapien
33	205	10.5	90	1 YAI1_CHICK	P12264 gallus gall

#### RESULT 1

ID	41_HUMAN	STANDARD;	PRT;	864 AA.
AC	P11171; P11176; Q9Y578; Q9Y579; Q14245;			
DT	01-JUL-1989 (Rel. 11, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	PROTEIN 4.1 (BAND 4.1) (P4.1) (EPB4.1).			
GN	EPB41 OR E41P.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (NON-ERYTHROID FORM).			
RX	MEDLINE=88234496; PubMed=3375238;			
RA	Tang T.K., Leto T.L., Correas I., Alonso M.A., Marchesi V.T.,			
RA	Benz E.J. Jr.;			
RT	"Selective expression of an erythroid-specific isoform of protein			
RT	4.1.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 85:3713-3717(1988).			
RN	[2]			
RP	SEQUENCE FROM N.A. (NON-ERYTHROID FORM).			
RX	MEDLINE=89132003; PubMed=3223413;			
RA	Tang T.K., Leto T.L., Marchesi V.T., Benz E.J. Jr.;			
RT	"Expression of specific isoforms of protein 4.1 in erythroid and non-			
RT	erythroid tissues.";			
RL	Adv. Exp. Med. Biol. 241:81-95(1988).			
RN	[3]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 378-393 (ERYTHROID ISOFORM).			
RC	TISSUE=Reticulocytes;			
RX	MEDLINE=87092279; PubMed=3467321;			
RA	Conboy J., Kan Y.W., Shohet S.B., Mohandas N.;			
RT	"Molecular cloning of protein 4.1, a major structural element of the			
RT	human erythrocyte membrane skeleton.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 83:9512-9516(1986).			
RN	[4]			
RP	SEQUENCE FROM N.A. (ISOFORM 3).			
RX	MEDLINE=91217063; PubMed=2022644;			
RA	Conboy J.G., Chan J.Y.C., Chasis J.A., Kan Y.W., Mohandas N.;			
RT	"Tissue- and development-specific alternative RNA splicing regulates			
RT	expression of multiple isoforms of erythroid membrane protein 4.1.";			
RL	J. Biol. Chem. 266:8273-8280(1991).			
RN	[5]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RA	Huang S.C., Wang C., Lichtenauer U., Vortmeyer A., Zhuang Z.;			
RT	"Sequence of protein 4.1 from a human neuroblastoma cell line: LAN5.";			
RT	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
RN	[6]			
RP	SEQUENCE OF 157-227 FROM N.A., AND VARIANT ILE-214.			
RA	Lichtenauer U., Huang S.C., Vortmeyer A., Zhuang Z.;			
RT	"Valine to isoleucine polymorphism in exon 4 of human protein 4.1.";			
RT	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
RN	[7]			
RP	SEQUENCE OF 648-714.			
RX	MEDLINE=87008553; PubMed=3531202;			

#### ALIGNMENTS

34	120	6.1	2492	1	TALA_DICDI	P54633 dictyosteli
35	119	6.1	2541	1	TALI_HUMAN	Q9Y490 homo sapien
36	116	5.9	2541	1	TALI_MOUSE	P26039 mus musculus
37	106	5.4	1429	1	EXPA_DROME	Q07436 drosophila
38	98.5	5.0	2215	1	MY7A_HUMAN	Q13402 homo sapien
39	96.5	4.9	2215	1	MY7A_MOUSE	P97479 mus musculus
40	94	4.8	488	1	NFSI_CANAL	P87185 candida alb
41	93	4.8	522	1	SNX1_HUMAN	Q13596 homo sapien
42	92.5	4.7	3635	1	LMA5_MOUSE	Q61001 mus musculus
43	91.5	4.7	663	1	TERM_ADEB3	O55439 bovine aden
44	91	4.7	484	1	NFS1_CANMA	P87187 candida mal
45	91	4.7	780	1	CUL3_CAEEL	Q17391 caenorhabdi



RT Identification, cloning, and expression of a cytosolic megakaryocyte  
 RT protein-tyrosine-phosphatase with sequence homology to cytoskeletal  
 RL protein 4.1.1.  
 RC Proc. Natl. Acad. Sci. U.S.A. 88:5867-5871(1991).  
 CC - FUNCTION: MAY ACT AT JUNCTIONS BETWEEN THE MEMBRANE AND THE  
 CC CYTOSKELETON.  
 CC - CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATASE + H(2)O =  
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.  
 CC - SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
 CC - SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.  
 CC - SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
 CC TYROSINE PHOSPHATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M68941; AAA36530.1; -;  
 CC PIR; A41105; A41105.  
 CC HSP; P18052; LYFO.  
 CC MIM; 176878; -;  
 CC InterPro; IPR000299; Band\_4.1.  
 CC InterPro; IPR001478; PDZ.  
 CC InterPro; IPR000387; TYR\_phosphatase.  
 CC InterPro; IPR000242; TYR\_prot\_phptase.  
 CC Pfam; PF00373; Band\_41; 1.  
 CC Pfam; PF00595; PDZ; 1.  
 CC Pfam; PF00102; Y\_phosphatase; 1.  
 CC PRINTS; PR00700; PRTYPHPTASE.  
 CC PRINTS; PR00935; BAND41.  
 CC SMART; SM00295; B41; 1.  
 CC SMART; SM00228; PDZ; 1.  
 CC SMART; SM00194; FPPC; 1.  
 CC PROSITE; PS00660; BAND\_41\_1; 1.  
 CC PROSITE; PS00661; BAND\_41\_2; 1.  
 CC PROSITE; PS50057; BAND\_41\_3; 1.  
 CC PROSITE; PS50106; PDZ; 1.  
 CC PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
 CC PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
 CC PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 CC Structural protein; Cytoskeleton; Hydrolase.  
 KW DOMAIN 84 241 BAND 4.1-LIKE.  
 FT DOMAIN 517 589 PDZ.  
 FT DOMAIN 677 926 PROTEIN-TYROSINE PHOSPHATASE.  
 FT ACT\_SITE 852 852 BY SIMILARITY.  
 FT SEQUENCE 926 AA; 105911 MW; 4DAC6A87A675CFB0 CRC64;  
 SQ  
 Query Match 34.7%; Score 678.5; DB 1; Length 926;  
 Best Local Similarity 41.1%; Pred. No. 1.7e-46;  
 Matches 146; Conservative 62; Mismatches 124; Indels 23; Gaps 6;  
 QY 35 PSGLKLSIK-----IQMLDDTQEAPEVQAPGKVLDDAVCNHNLVVEG 78  
 Db 8 PAGRINYVASELARDROHTEVVCNILLNDNTVQAEKVNKHQDQGVLLDVVFKHLDLTDQ 67  
 QY 79 DYFGLFPDHHKITV-WDLLLKPVKQIRPKHVVFVVKVFFPPDHTQLQELTRYLFA 137  
 Db 68 DYFGLQADDSDNPRWLDNPNKPIRQLKRGSPYSLNFRVKFVSPNKLQBEYTRYQYF 127  
 QY 138 LQVKODLAQGLTCDNDSALISHTVQSEIGDFDEALDRE-HLAKNKYIPQO-DALEDK 195  
 Db 128 LQIKODILTRULPCPSNTAALLASFAVQSELGSDYDQSENLSGLSDYSFIPNQPODFEKE 187  
 QY 196 IVEFHNHNGQTPAESDFOLLETARLEMYGIRLHPAKDREGTKINLAVANTGILVFOGF 255  
 Db 188 IAKLHQHGLGLSPAEEFNLYNTARTLEYGVFFHVARQDSNNEIMVSGGILLYKNR 247  
 QY 256 TKINAFNNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHFAF 315

Db 248 VRNTPFPWKIVKISFKCKQFFIQLRKELHESRETLLGFNMVNYRACKNLWKACVHHFF 307  
 QY 316 FRLFEPPKPKPVL---FSRGSFRFSRGTQKQVLDYVKEGHHKKVQFERKHSK 367  
 Db 308 FRL-DRPLPPQKNFFAHYFTLGSKFYCGRTEVSQVQYKGRKANKDRVFARSPSK 361  
 RESULT 3  
 EMBL\_HUMAN STANDARD; PRT; 881 AA.  
 AC Q9HAG0; O15046;  
 DT 20-AUG-2001 (Rel. 40, Created)  
 DT 20-AUG-2001 (Rel. 40, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE BAND 4.1-LIKE PROTEIN 1.  
 GN EPB41L1 OR KIAA0338.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=97349984; PubMed=9205841;  
 RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.;  
 RN DNA Res. 4:141-150(1997).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA Tromans A.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 CC - SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AB002336; BAA20796.1; ALT\_INIT.  
 CC EMBL; AL121895; CAC09920.1; -;  
 CC MIM; 602879; -;  
 CC InterPro; IPR000299; Band\_4.1.  
 CC Pfam; PF00373; Band\_41; 1.  
 CC PRINTS; PR00935; BAND41.  
 CC SMART; SM00295; B41; 1.  
 CC PROSITE; PS00660; BAND\_41\_1; 1.  
 CC PROSITE; PS00661; BAND\_41\_2; 1.  
 CC PROSITE; PS50057; BAND\_41\_3; 1.  
 KW Cytoskeleton.  
 FT DOMAIN 94 308 BAND 4.1-LIKE.  
 FT CONFLICT 484 484 MISSING (IN REF. 2).  
 FT CONFLICT 729 729 MISSING (IN REF. 2).  
 FT SEQUENCE 881 AA; 98502 MW; D923CF554EDB41D3 CRC64;  
 SQ  
 Query Match 34.2%; Score 669.5; DB 1; Length 881;  
 Best Local Similarity 41.5%; Pred. No. 8.4e-46;  
 Matches 153; Conservative 59; Mismatches 138; Indels 19; Gaps 10;  
 QY 18 ENSGIS---TLERGQKPPPTPSKLVLS--IKIQMLDDTQEAPEVQAPGKVLDDAVCNH 72  
 Db 70 EADGLSERTTPSKAQSPQKIAKKYKSAICRVTLDDASEYEVEKHGRGVFLDLVCEH 129  
 QY 73 LNLVEGDYGLGLEPDKHKITVWLDLLKPIVKQIR-RPKHVVFVVKVFFPPDHTQLQEEL 131  
 Db 130 LNLLEKDYGLTFCDDADSOKNWLDPDSKEIKKQIRSSPNFA--FTVKFYPPDPAQLTEDI 187



CC -1- TISSUE SPECIFICITY: FOUND EXCLUSIVELY IN PHOTORECEPTORS FOLLOWING  
 CC THE TERMINAL MITOSIS OF RETINAL NEURONS. WHEN RETINAL  
 CC SYNTOGENESIS IS COMPLETE, PROTEIN 4.1 IS ALSO EXPRESSED IN THE  
 CC INNER RETINA. IN ADULT AMPHIBIAN RETINAS, PROTEIN 4.1 IS DETECTED  
 CC IN PHOTORECEPTORS, BIPOLAR CELLS, AND GANGLION CELL AXONS.  
 CC -1- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES  
 CC AND EACH PHOSPHORYLATION EVENT SELECTIVELY MODULATES THE PROTEIN'S  
 CC FUNCTIONS.  
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M20621; AAA49695.1; --  
 CC PIR: A29901; A29901.  
 CC PIR: A37353; A37353.  
 CC InterPro: IPR000299; Band\_4.1.  
 CC Pfam: PF00373; Band\_41; 1.  
 CC PRINTS: PR00935; BAND41.  
 CC SMART: SM00295; B41; 1. 1.  
 CC PROSITE: PS00660; BAND\_41\_1; 1.  
 CC PROSITE: PS00661; BAND\_41\_2; 1.  
 CC PROSITE: PS50057; BAND\_41\_3; 1.  
 CC Structural protein; Cytoskeleton; Phosphorylation.  
 CC KW DOMAIN 247 403 BAND 4.1-LIKE.  
 CC FT SEQUENCE 801 AA; 89429 MW; 07FA508552359A0F CRC64;  
 CC -----  
 CC Query Match 33.6%; Score 658; DB 1; Length 801;  
 CC Best Local Similarity 41.5%; Pred. No. 6.2e-45;  
 CC Matches 139; Conservative 60; Mismatches 116; Indels 20; Gaps 6;  
 QY 28 GQKPPPTSGKL-----VSIKQMLDDTQAEFVQAPGKVLDDAVCNHNL 75  
 DB 169 GKEPIKPEGESKASHKVVRRSPNCRKVTLLDDTVYECDEKHKAGQDIFKVCVSHLNI 228  
 QY 76 VEGDYFGLFEPPDHKKITVMDLLKPTVKOIRPKPHVVKFV--VKFPPDHTOLOQELTR 133  
 DB 229 VEEDYFGLAIWESPCTKVMPLDKTRKOVHGGP---CEFTSNVKYFPDPAQLSDDIR 285  
 QY 134 YLFALQVKQDLAQRGRLTCDNTSAALLSHIVQSEIGDFDEALDR-EHLAKNKYIPQODA- 191  
 DB 286 YLCLQLKRLDIFSGRLPCSFATLALLGTYTVQSEVGDYEDLHGVDYVSEFKLSPNQTKD 345  
 QY 192 LEDKIVFHHNHGQTPASDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGILV 251  
 DB 346 LEEKVGLHKSYSRMTPAQADLEFLENAKKLTWYGVDIHQAKDLEGVDIKLGVCSGLMV 405  
 QY 252 FQGTETKINAFNAVKRLFKRKRFLKLRDANSAYQDTLEFLMASRDFCKSFWKICVE 311  
 DB 406 FKDNLRINRPWPVKLVKISYKRSRFFIKIRPGEOQVESTIGFKLPSYKRAKKLWKCVE 465  
 QY 312 HHAFRLFEPPKPKPKVLFVSRGSSFRSGRTQKQ 346  
 DB 466 HHTFRL-TSTESIPKHFRLSLGSLTFRYSGRTOAQ 499

## RESULT 6

PTP1\_CAEEL STANDARD; PRT; 1026 AA.  
 AC P28191; P45449;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROTEIN-TYROSINE PHOSPHATASE 1 (EC 3.1.3.48).  
 OS PTP-1 OR C48D5.2.  
 GN Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Lightning J., Thomas K.;  
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RX SEQUENCE OF 844-950 FROM N.A.  
 RX MEDLINE=91139172; PubMed=1704870;  
 RT "Protein tyrosine phosphatase domains from the protochordate Styela  
 RT plicata.";  
 RL Immunogenetics 33:33-41(1991).  
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =  
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.  
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
 CC TYROSINE PHOSPHATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: Z36237; CAA85272.1; --  
 CC EMBL: Z48241; CAA85272.1; JOINED.  
 CC EMBL: Z48241; CAA88287.1; --  
 CC EMBL: Z36237; CAA88287.1; JOINED.  
 CC EMBL: M38013; AAA28127.1; --  
 CC HSSP: P28827; IRPM.  
 CC WormPep: C48D5.2; CE17578.  
 CC InterPro: IPR000299; Band\_4.1.  
 CC InterPro: IPR001478; PDZ.  
 CC InterPro: IPR000387; Tyr\_phosphatase.  
 CC InterPro: IPR000242; Tyr\_prot\_phptase.  
 CC Pfam: PF00373; Band\_41; 1.  
 CC Pfam: PF00595; PDZ; 1.  
 CC Pfam: PF00102; Y\_phosphatase; 1.  
 CC PRINTS: PR00700; PRTYPHPTASE.  
 CC PRINTS: PR00935; BAND41.  
 CC SMART: SM00295; B41; 1.  
 CC SMART: SM00228; PDZ; 1.  
 CC SMART: SM00194; PTPC; 1.  
 CC PROSITE: PS00660; BAND\_41\_1; 1.  
 CC PROSITE: PS00661; BAND\_41\_2; 1.  
 CC PROSITE: PS00057; BAND\_41\_3; 1.  
 CC PROSITE: PS50106; PDZ; 1.  
 CC PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 1.  
 CC PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 1.  
 CC PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 CC Hydrolase; Cytoskeleton.  
 CC KW DOMAIN 87 245 BAND 4.1-LIKE.  
 CC FT DOMAIN 777 1026 PROTEIN-TYROSINE PHOSPHATASE.  
 CC FT ACT\_SITE 952 952 BY SIMILARITY.  
 CC SEQUENCE 1026 AA; 115093 MW; 2F43F7A614EDBC59 CRC64;  
 CC -----  
 CC Query Match 31.2%; Score 610; DB 1; Length 1026;  
 CC Best Local Similarity 39.1%; Pred. No. 6e-41;  
 CC Matches 147; Conservative 60; Mismatches 145; Indels 24; Gaps 10;

QY 13 RLGAPENSGISTLERGOKPPTPSGKLVSTKIOMLDDTQAEFVQAPGKVLDDAVCNH 72  
 DB 2 RLGSNDYVQRTAIGTQTPVKTPPPNQIRCTVTFLOSTSYHFEIERNLSGIVLLEKFN 61  
 QY 73 LNLVEGDYFGLFE-----PDHKKITVMDLLKPTVKQ-IRRPKHVVKVVFPPDH 124  
 DB 62 LEIEXDYFGLVFIADVNSAQQK---WLDPSKLNLRKQMICPPYHLF--FRVKEYVRDP 116





[illegible]



Db 134 IAAQCAVAIEGLDHDYKTAGVSYREVPQKEELEAEIERIHKTLTMCQAPSEAE 193  
QY 213 FOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFOGFTTKINAFNNAKVRKLSFK 272  
Db 194 LNYLRTAKSLEMYGVDLHPVYGENKSEYFLGLTPSGVVVYKKNKQGVYFWPRITKVHFK 253  
QY 273 RKRFLIK-LRPDANSAYQDTLEFLMASRDFCKSFVKICVEHHAFFRLFEPPKPKPVLF 331  
Db 254 EQFELRVLGKDCN---ETSFEEARSKTACKHLWKCSVEHHTTFRMPDPTESNSLSRKL 310  
QY 332 SRGS-SFREGSGRTQKV 347  
Db 311 KFGSISYKHRYTALQM 327

## RESULT 12

PTNL\_RAT PTNL\_RAT STANDARD; PRT; 1175 AA.  
AC Q62728; Q62732; PTNL\_RAT STANDARD; PRT; 1175 AA.  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 21 (EC 3.1.3.48)  
GN PTPN21 OR PTP2E.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY;  
RX MEDLINE=95104449; PubMed=7805871;  
RA L'Abbe D., Banville D., Tong Y., Stocco R., Masson S., Ma S.,  
RA Fantus G., Shen S.H.;  
RT Identification of a novel protein tyrosine phosphatase with sequence  
RT homology to the cytoskeletal proteins of the band 4.1 family.;  
RL FEBS Lett. 356:351-356(1994).  
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2O) =  
CC PROTEIN TYROSINE + PHOSPHATE.  
CC -!- TISSUE SPECIFICITY: PARTICULARLY ABUNDANTLY IN ADRENAL GLANDS.  
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
CC TYROSINE PHOSPHATASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; U17971; AAA62153.1; -;  
DR EMBL; U18293; AAA62154.1; -;  
DR HSSP; P18031; 1PTT.  
DR InterPro; IPR000299; Band.4.1.  
DR InterPro; IPR000387; TYR phosphatase.  
DR InterPro; IPR000242; Tyr\_prot\_phphatase.  
DR Pfam; PF00373; Band.41; 1.  
DR Pfam; PF01102; Y-phosphatase; 1.  
DR PRINTS; PR00700; PRTYPHPHTASE.  
DR PRINTS; PR00935; BAND41.  
DR SMART; SM00295; B41; 1.  
DR SMART; SM00194; PTPC; 1.  
DR PROSITE; PS00860; BAND\_41\_1; 1.  
DR PROSITE; PS00661; BAND\_41\_2; 1.  
DR PROSITE; PS00057; BAND\_41\_3; 1.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 1.  
DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 1.  
KW Structural protein; Cytoskeleton; Hydrolase; Alternative splicing.

FT DOMAIN 77 241 BAND 4.1-LIKE.  
FT DOMAIN 922 1175 PROTEIN-TYROSINE PHOSPHATASE.  
FT ACT\_SITE 1109 1109 BY SIMILARITY.  
FT VARSPLIC 1 839 MISSING (IN ISOFORM 2E).  
SQ SEQUENCE 1175 AA; 133411 MW; 82A684F1C0F5ECF7 CRC64;  
Query Match 18.8%; Score 368; DB 1; Length 1175;  
Best Local Similarity 30.8%; Pred. No. 1.6e-21;  
Matches 102; Conservative 65; Mismatches 140; Indels 24; Gaps 10;  
QY 25 LERGOKPPPTSGKLVSIKIOMLDTDQEAPEVFPQAPGKVLDDAVCNHNLNVEGDYFGL 84  
Db 9 LKTRRYTVSSKSLVA-RIQLNNEFEVETLSVESTGQESLEAVAQRLREITVYFSLW 67  
QY 85 FPDHKKITVMDLLKPIVKQIRPKHV---VKFVVVFFPPDHTQOEELRYLFAQVK 141  
Db 68 YYNKQNRWRWDLKPKQL--DKHALEPTVYFVGVVPSVQLQEQETTRYQYVQLK 125  
QY 142 QDLAQRLTCNDTSRAALLSHIVQSEIGDFDEALDREHLAKNKYIP-----QDALED-- 194  
Db 126 KDLVEGNLPTLEQAIQLAGLAVQADFGDFDQYESQDFLOKTFALLVGVWLQDEKLEEA 185  
QY 195 -KIVEFHHNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGILV-- 251  
Db 186 QKVALHQQYRGITAPEAEMLYMQEVRMDGGEESYPKADSGQSDISIGACLDGIFVKH 245  
QY 252 FOGFTKINAFNNAKVRKLSFKRFLIKRLPDANSAYQDTLEFLMASRDFCKSFNKCIVE 311  
Db 246 KNGRPVP-VFRWHDIANMSHKSFPALEL--ANK--EETIQFTEDMETAKYVWRLCVA 299  
QY 312 HHAFRLFEPPKPKPVLF--RGSSFRFS 340  
Db 300 RHKFRVLRNQCNLQTOAATLNSVRRGSSRMS 330

## RESULT 13

PTNL\_MOUSE PTNL\_MOUSE STANDARD; PRT; 1176 AA.  
AC Q62136;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 21 (EC 3.1.3.48)  
DE (PROTEIN-TYROSINE PHOSPHATASE PTP-RL10).  
GN PTPN21.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6; TISSUE=Liver;  
RX MEDLINE=95140431; PubMed=7838537;  
RA Higashitsuji H., Aril S., Furutani M., Inamura M., Kaneko Y.,  
RA Takenawa J., Nakayama H., Fujita J.;  
RT "Enhanced expression of multiple protein tyrosine phosphatases in the  
RT regenerating mouse liver: Isolation of ppp-RL10, a novel cytoplasmic-  
RT type phosphatase with sequence homology to cytoskeletal protein  
RT 4.1.";  
RL Oncogene 10:407-414(1995).  
CC -!- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF GROWTH AND  
CC DIFFERENTIATION OF LIVER CELLS.  
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2O) =  
CC PROTEIN TYROSINE + PHOSPHATE.  
CC -!- TISSUE SPECIFICITY: LIVER.  
CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
CC TYROSINE PHOSPHATASE FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its



```
QY 319 FEEPKPKPVLFSGSSFRSRTQKQVLDYVKEGGHKKVQF-ERKHSKI 368
Db 292 ---RRRKPDTIEVQQMKAQAREKHOKQERALLENEKKKKRLAEAKEKEKI 339

RESULT 15
MOES_HUMAN
ID MOES_HUMAN STANDARD; PRT; 576 AA.
AC P26038;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MOESIN (MEMBRANE-ORGANIZING EXTENSION SPIKE PROTEIN).
GN MSN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-15; 53-59 AND 413-434.
RC TISSUE=Placenta;
RX MEDLINE=92020840; PubMed=1924289;
RA Lankes W.T., Furthmayr H.;
RT "Moesin: a member of the protein 4.1-talin-ezrin family of proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:8297-8301(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA Howden P.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC !- FUNCTION: PROBABLY INVOLVED IN CONNECTIONS OF MAJOR CYTOSKELETAL
CC STRUCTURES TO THE PLASMA MEMBRANE.
CC !- TISSUE SPECIFICITY: IN ALL TISSUES AND CULTURED CELLS STUDIED.
CC !- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M69066; AAA36322.1; -
CC EMBL; Z98946; CAB46379.1; -
CC PIR; A41289; A41289.
CC Aarhus/Ghent-2DPAGE; 3515; IEF.
CC Aarhus/Ghent-2DPAGE; 3516; IEF.
CC MIM; 309845; -
CC InterPro; IPR000299; Band_4.1.
CC InterPro; IPR000798; Ezrin_radixin_moesin.
CC Pfam; PF00373; Band_41; 1.
CC Pfam; PF00769; ERM; 1.
CC PRINTS; PR00661; ERMFAMILY.
CC PRINTS; PR00935; BAND41.
CC SMART; SM00295; B41; 1.
CC PROSITE; PS00660; BAND_41_1; 1.
CC PROSITE; PS00661; BAND_41_2; 1.
CC PROSITE; PS00507; BAND_41_3; 1.
CC Structural protein; Cytoskeleton.
CC INIT MET 0
FT DOMAIN 57 224 BAND 4.1-LIKE.
SQ SEQUENCE 576 AA; 67689 MW; 96CE7C28D55A8176 CRC64;
```

```
Query Match 18.5%; Score 362.5; DB 1; Length 576;
Best Local Similarity 27.9%; Pred No. 1.7e-21;
Matches 98; Conservative 77; Mismatches 143; Indels 33; Gaps 10;

QY 38 KLVSIKIQMLDQTQEAFFVQRPAGKVLNHLNVLVEGDYFGLEFPDHHKITYWLDL 97
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 2 KTISVRVTTMDAELE-FAIQPNTTGKQLFDQVVKVTIGLREVWFFGLQYQDTKGFTWLKL 60

QY 98 LRPVVKQ-IRRPKHVVVVKFVFFPPD-HTQLQELTRYLFALQVKQDLAQGRLCNDTS 155
```

```
Db 61 NKKVTAQDVKRESPLLFKFRAFYEDVSEELIQDITQRLFFLOVKEGILNDDIYCPTT 120
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 156 AALLISHIVQSEIGDFDEALDRE-HLAKNKYIPQ-----QDALEDKIYVEFHNNHIG 205
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 121 AVLLASYAVQSKYGVDFNKVKHSGYLAGDKLLPQRVLEQHKLNKQDWEERIQVWHEEHRG 180
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 206 QTPASDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILYVFOGFTKIN---AFN 262
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 181 MLREDAVLEYLKIAODLEMYGVNYFSIKNKGSELWGLVDALGLNIYEQNDRLTPKIGFP 240
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 263 WAKVRKLSEFKRRFLI----KLRPDANSAYQDTLEFLMASRDFCKSFWRICVVEHHAFRL 318
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 241 WSEIRNISFNKKFVYKPIDKKAPD-----FVYAPRLRINKRILALCMGNHLYW- 291
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
QY 319 FEEPKPKPVLFSGSSFRSRTQKQVLDYVKEGGHKKVQF-ERKHSKI 368
| : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 292 ---RRRKPDTIEVQQMKAQAREKHOKQERALLENEKKKKREMAEKEKEKI 339
```

Search completed: December 6, 2001, 08:53:08  
Job time: 380 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5

Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:54:30 ; Search time 77.11 Seconds  
(without alignments)  
709.452 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_1\_374

Perfect score: 1956

Sequence: 1 MGTEQRPTGSLGAPENS.....GHKKVQPERKXKIHRSRL 374

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTEMBL\_17.\*

1: sp-archaea.\*

2: sp-bacteria.\*

3: sp-fungi.\*

4: sp-human.\*

5: sp\_invertebrate.\*

6: sp\_mammal.\*

7: sp\_mhc.\*

8: sp-organelle.\*

9: sp-phage.\*

10: sp-plant.\*

11: sp-rodent.\*

12: sp-virus.\*

13: sp-vertebrate.\*

14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1956	100.0	1045	Q9Y4F1	Q9y4f1 homo sapien
2	1246.5	63.7	1054	Q94887	Q94887 homo sapien
3	867	44.3	330	Q9VN68	Q9vn68 drosophila
4	802.5	41.0	1111	O17905	O17905 caenorhabdi
5	699.5	35.8	988	O170318	O70318 mus musculu
6	691.5	35.4	929	Q9WV92	Q9wv92 mus musculu
7	691.5	35.4	963	Q9JMB2	Q9jmb2 rattus norv
8	691.5	35.4	1105	Q9JMB3	Q9jmb3 rattus norv
9	690.5	35.3	865	Q9BRP5	Q9brp5 homo sapien
10	690.5	35.3	1115	Q9Y2J2	Q9y2j2 homo sapien
11	689.5	35.3	1005	O43491	O43491 homo sapien
12	684	35.0	503	Q95713	Q95713 homo sapien
13	681.5	34.8	864	Q9Y579	Q9y579 homo sapien
14	677	34.6	706	Q9R102	Q9r102 mus musculu
15	671.5	34.3	617	Q9N179	Q9n179 bos taurus
16	670.5	34.3	641	Q14245	Q14245 homo sapien
17	669.5	34.2	879	O9H4G0	O9h4g0 homo sapien
18	669.5	34.2	934	O15046	O15046 homo sapien
19	667.5	34.1	879	Q9Z2H5	Q9z2h5 mus musculu

20	666.5	34.1	879	11	Q9WTP0	Q9wtp0 rattus norv
21	666.5	34.1	1551	11	Q9WTP1	Q9wtp1 rattus norv
22	662.5	33.9	926	11	Q9WU22	Q9wu22 mus musculu
23	661.5	33.8	703	5	Q9V8S0	Q9v8s0 drosophila
24	661.5	33.8	889	5	Q9V8R8	Q9v8r8 drosophila
25	661.5	33.8	1698	5	O24440	O24440 drosophila
26	661.5	33.8	1698	5	Q9V8R9	Q9v8r9 drosophila
27	589.5	30.1	505	6	Q9MRU8	Q9myu8 canis famil
28	587.5	30.0	687	4	Q9H975	Q9h975 homo sapien
29	584.5	29.9	504	11	Q99KZ8	Q99kz8 mus musculu
30	567	29.0	972	5	Q9VFU8	Q9vfu8 drosophila
31	552.5	28.2	527	11	Q9JMC8	Q9jmc8 mus musculu
32	542.5	27.7	504	4	Q9H328	Q9h328 homo sapien
33	542.5	27.7	518	4	Q9P2V3	Q9p2v3 homo sapien
34	542.5	27.7	913	4	Q9H329	Q9h329 homo sapien
35	531.5	27.2	987	5	Q9NDP4	Q9ndp4 ciona intes
36	527	26.9	568	5	Q9GP93	Q9gp93 caenorhabdi
37	524	26.8	283	4	Q9BQV2	Q9bvq2 homo sapien
38	514.5	26.3	440	4	Q9NX84	Q9nx84 homo sapien
39	480.5	24.6	598	4	Q9HCS5	Q9hcs5 homo sapien
40	438.5	22.4	916	5	Q9W0R3	Q9w0r3 drosophila
41	434.5	22.2	1199	5	Q9W9Y3	Q9w9y3 drosophila
42	423	21.6	639	5	Q9XV97	Q9xv97 caenorhabdi
43	423	21.6	4667	5	Q9TVI9	Q9tv19 caenorhabdi
44	416.5	21.3	572	5	Q9VKY7	Q9vky7 drosophila
45	363.5	18.6	577	11	O35763	O35763 rattus norv

## ALIGNMENTS

RESULT 1

Q9Y4F1  
ID Q9Y4F1 PRELIMINARY; PRT: 1045 AA.  
AC Q9Y4F1;  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE CDEP.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_Taxid=9606;  
RN [1]

RP SEQUENCE FROM N.A.  
RC TISSUE=CARTILAGE;  
RX MEDLINE=98086358; PubMed=9425278;

RA Koyano Y., Kawamoto T., Shen M., Yan W., Noshiro M., Fujii K.,  
RA kato Y.;

RT "Molecular cloning and characterization of CDEP, a novel human protein  
containing the ezrin-like domain of the band 4.1 superfamily and the  
Dbl homology domain of Rho guanine nucleotide exchange factors.";  
RL Biochem. Biophys. Res. Commun. 241:369-375(1997).  
DR EMBL; AB008430; BAA24267.1; .  
DR HSSP; P08567; IPLS.  
DR InterPro; IPR000299; Band\_4.1.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR000219; RhoGEF.  
DR Pfam; PF00373; Band\_41; 1.  
DR Pfam; PF00169; PH; 2.  
DR Pfam; PF00621; RhoGEF; 1.  
DR PRINTS; PR00935; BAND41.  
DR SMART; SM00295; B41; 1.  
DR SMART; SM00233; PH; 2.  
DR SMART; SM00325; RhoGEF; 1.  
DR PROSITE; PS00660; BAND\_41\_1; UNKNOWN\_1.  
DR PROSITE; PS50057; BAND\_41\_3; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 2.  
SQ SEQUENCE 1045 AA; 118632 MW; 0E8B2D61C0F58417 CRC64;

Query Match 100.0%; Score 1956; DB 4; Length 1045;  
Best Local Similarity 100.0%; Pred. No. 1.6e-148;

Matches 374; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGEIEQRTPGSRGAPENSIGSTLERGQKPPPTPSGKLVSIKIOMLDDTQEAFAFVQRA 60  
 DB 1 MGEIEQRTPGSRGAPENSIGSTLERGQKPPPTPSGKLVSIKIOMLDDTQEAFAFVQRA 60

QY 61 PGKVLDAVNCNHLNVEGDYFGLFDPHKKITVWDLKKPIVQIRRPKHVVVKKFF 120  
 DB 61 PGKVLDAVNCNHLNVEGDYFGLFDPHKKITVWDLKKPIVQIRRPKHVVVKKFF 120

QY 121 PPDHTQLQELTRYLFALQVQDLAOGRLTNCNDSALSHVQSEIGDFDEALDREHL 180  
 DB 121 PPDHTQLQELTRYLFALQVQDLAOGRLTNCNDSALSHVQSEIGDFDEALDREHL 180

QY 181 AKNKYIPQDALEDKIVEFHNNHIGOTPAESDFQLEIARRLMYGIRLHPAKDREGTKI 240  
 DB 181 AKNKYIPQDALEDKIVEFHNNHIGOTPAESDFQLEIARRLMYGIRLHPAKDREGTKI 240

QY 241 NLAVANTGILVFGFTKINAFNNAKVRKLSFKRKRFLIKLRPDANSAYQDTLEFLMASRD 300  
 DB 241 NLAVANTGILVFGFTKINAFNNAKVRKLSFKRKRFLIKLRPDANSAYQDTLEFLMASRD 300

QY 301 FCKSEFKICVEHHAFFLPEEPKPKPVLFSGSSFRSGRTQKQVLDYVKEGGHKVQ 360  
 DB 301 FCKSEFKICVEHHAFFLPEEPKPKPVLFSGSSFRSGRTQKQVLDYVKEGGHKVQ 360

QY 361 FERKSHKHSIRSL 374  
 DB 361 FERKSHKHSIRSL 374

RESULT 2

ID O94887 PRELIMINARY; PRT; 1054 AA.

AC O94887;

DT 01-MAY-1999 (TREMBlrel. 10, Created)

DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

DE K1AA0793 PROTEIN.

GN K1AA0793.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=BRAIN;

RX MEDLINE=99087487; PubMed=9872452;

RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;

RA "Prediction of the coding sequences of unidentified human genes. XI. The complete sequences of 100 new cDNA clones from brain which code for large proteins in vitro.";

RT DNA Res. 5:277-286(1998).

RL [2]

RN SEQUENCE OF 1-754 FROM N.A.

RP MEDLINE=99063792; PubMed=9847074;

RA Sulston J.E., Waterston R.;

RA "Toward a complete human genome sequence.";

RT Genome Res. 8:1097-1108(1998).

RN [3]

RN SEQUENCE OF 1-754 FROM N.A.

RA Sun H., Stoneking T., Langston Y., LaPlant Y.;

RA "The sequence of Homo sapiens BAC clone RC442F18.";

RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [4]

RN SEQUENCE OF 1-754 FROM N.A.

RA Waterston R.H.;

RA Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

RN [5]

RN SEQUENCE OF 1-754 FROM N.A.

RA Waterston R.;

RA Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.

[6]

RN SEQUENCE OF 1-754 FROM N.A.

RP Waterston R.;

RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB018336; BAA34513.1; -

DR EMBL; AC005104; AAD12224.1; -

DR HSSP; P08567; IPLS.

DR InterPro; IPR000299; Band\_4.1.1.

DR InterPro; IPR001849; PH.

DR InterPro; IPR000219; RhoGEF.

DR Pfam; PF00373; Band\_41; 1.

DR Pfam; PF00169; PH; 2.

DR Pfam; PF00621; RhoGEF; 1.

DR PRINTS; PR00935; BAND41.

DR SMART; SM00295; B41; 1.

DR SMART; SM00233; PH; 2.

DR SMART; SM00325; RhoGEF; 1.

DR PROSITE; PS00660; BAND\_41\_1; UNKNOWN\_1.

DR PROSITE; PS50057; BAND\_41\_3; 1.

DR PROSITE; PS50003; PH\_DOMAIN; 2.

SQ SEQUENCE 1054 AA; 119887 MW; 7DC279F69A307E5A CRC64;

Query Match 63.7%; Score 1246.5; DB 4; Length 1054;  
 Best Local Similarity 62.5%; Pred. No. 1.5e-91;  
 Matches 237; Conservative 58; Mismatches 79; Indels 5; Gaps 2;

QY 1 MGEIEQR---PTPGSRGAPENSIGSTLERGQKPPPTPSGKLVSIKIOMLDDTQEAFAFV 56  
 DB 1 MGEIEGYRVLQTAGMRLGAQTPTGVSTLPGQTLLRPMQEKHLHLVKLLDNTMEIFDI 60

QY 57 PQRAPGKVLDAVNCNHLNVEGDYFGLFDPHKKITVWDLKKPIVQIRRPKHVVVKKFV 116  
 DB 61 EPKCDGQVLLTQVWKRLNLVECDYFGMEFQNTQSYWTLPEPKPIIRQIRRPKNVVLRLA 120

QY 117 VKFFPDHTQLQELTRYLFALQVQDLAOGRLTNCNDSALSHVQSEIGDFDEALD 176  
 DB 121 VKFFPDGQLEETRYLFALQVQDLAOGRLTNCNDSALSHVQSEIGDFDEALD 180

QY 177 REHLAKNKYIPODALEDKIVEFHNNHIGOTPAESDFQLEIARRLMYGIRLHPAKDRE 236  
 DB 181 REHLKVNLYLPGQHQHLEKLEFQHQVGTQPAESDFQLEIARRLMYGIRLHPAKDRE 240

QY 237 GTKINLAVANTGILVFGFTKINAFNNAKVRKLSFKRKRFLIKLRPDANSAYQDTLEFLM 296  
 DB 241 GTKIQAVSHMGVLVFGFTKINTFNWSKVRKLSFKRKRFLIKLRHPVHGYPQDTLEFL 300

QY 297 ASRDFCKSWKICVEHHAFFLPEEPKPKPVLFSGSSFRSGRTQKQVLDYVKEGGH 356  
 DB 301 GSRDECKNFKWICVEYTFEFLDQPKAKAVFFSGSSFRSGRTQKQVLDYVKEGGH 360

QY 357 KKVOFERKSHKHSIRSL 374  
 DB 361 KRPIYERRHSHKHSIRSL 379

RESULT 3

O9VN68 PRELIMINARY; PRT; 330 AA.

ID O9VN68

AC O9VN68;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DE CG1283 PROTEIN.

GN CG1283.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;



RX MEDLINE-20196006; PubMed-10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Miklos G.L.G.,  
RA Ballwe R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
RA Jalali M., Kalush F., Karp G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., Mcintosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J.C., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
RT "The genome sequence of *Drosophila melanogaster*."  
RL Science 287:2185-2195 (2000).  
DR EMBL; AF003604; AAF52081.1; .  
DR FlyBase; FBgn0037286; CG1283.  
DR InterPro; IPR000299; Band\_4.1.  
DR Pfam; PF00373; Band\_41; 1.  
DR PRINTS; PR00935; BAND41.  
DR SMART; SM00295; B41; 1.  
DR PROSITE; PS00660; BAND\_41\_1; 1.  
DR PROSITE; PS00057; BAND\_41\_3; 1.  
DR PROSITE; PS00057; BAND\_41\_3; 1.  
SQ SEQUENCE 330 AA; 38707 MW; FBDAB30DBC86CBF CRC64;

Query Match 44.3%; Score 867; DB 5; Length 330;  
Best Local Similarity 51.5%; Pred. No. 9.2e-62;  
Matches 168; Conservative 64; Mismatches 84; Indels 10; Gaps 6;

QY 46 MLDSTQAEVQAPGKVLDAVCHNLVSGDYGLEFPD---HKKTIVWLDDLLKPIV 102  
DB 1 MLDSTMTFQVQAKALGRVLFQVQCQLNLEADYEGLEYQVSTHTK--YWLDEKPN 58  
QY 103 KQLRRPK-HVVYKFFVPPDHTLOEELRYLFLALQVKQDLAOCRLTCNDTSAALLIS 161  
DB 59 RQGLSLDPLVLCFCIKFTTPDPAQLLEEYTRYLCQKRLDRLTSLGLOCNALMAS 118  
QY 162 HIYQSGIGDF--DEALDREHLAKNKYIPOQDA--LEDKIVEFHHNHIGQTPAESDFOLLEI 218  
DB 119 YIVQASCGDFVDPDYDHTYLLSYRPVQNDATMQRKIMENHKHKGVSQPAEADNLLET 178  
QY 219 ARLEMYGRLHPADKREGTKINLAVANTGILVFGQFTKINAFNNAKVRKLSFKRRFLI 278  
DB 179 ARCELYGKMHKPAKDVGVPLNLAHMGITVFNQITRINTFSWAKIRKISFKRRFLV 238  
QY 279 KLRPDANSAYQDTLEFLMASRDFCKSFVKICVHHAFERLFE-EPKPKPKVLFSGSSF 337  
DB 239 KLPGEYGYKDTVEFFFGRECKNFWKCKVENHGHFFRCTAVQNTPRKTRVLSRGSF 298

QY 338 RFSGRTQKQVLDYVKEGGHKKVQFER 363  
DB 299 RYSGTKQKQIEFVRENVYKQNFQR 324  
RESULT 4  
O17905 PRELIMINARY; PRT; 1111 AA.  
ID O17905;  
AC O17905;  
DT 01-JAN-1998 (TREMBLrel. 05, Created)  
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE H05G16.1 PROTEIN.  
GN H05G16.1.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OC NCBI\_TaxID=6239;  
CX [1]  
RN White S.;  
RP Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen R.,  
RA Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,  
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkison-Sproat J., Wohldman P.;  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
elegans";  
RL Nature 368:32-38 (1994).  
DR EMBL; Z97190; CAB10024.1; .  
DR InterPro; IPR000299; Band\_4.1.  
DR Pfam; PF00373; Band\_41; 1.  
DR PRINTS; PR00935; BAND41.  
DR PROSITE; PS00660; BAND\_41\_1; UNKNOWN\_1.  
DR PROSITE; PS00057; BAND\_41\_3; 1.  
DR SMART; SM00295; B41; 1.  
SQ SEQUENCE 1111 AA; 124808 MW; 1EE53EB512C0FE2A CRC64;

Query Match 41.0%; Score 802.5; DB 5; Length 1111;  
Best Local Similarity 44.2%; Pred. No. 7.1e-56;  
Matches 165; Conservative 64; Mismatches 113; Indels 31; Gaps 6;

QY 22 ISTLERGQPPPTPS-----GKLSVIKIQMLDDTQEAPE----- 55  
DB 1 MSNIPRGVAGPPGPNSSKRGRLVCIKVRMLDDTVAVFHLGVICFFCVSPHGTYVFL 60  
QY 56 VQRAPGKVLDAVCHNLVSGDYGLEFPDHTLOEELRYLFLALQVKQDLAOCRLTCNDTSAALLIS 114  
DB 61 LQKALGQTLLEDCVCHNLLEDCDYGLSFIDINGNHCWLDREKTLRQITNGSTDAKY 120  
QY 115 FVYKFPDHTLOEELRYLFLALQVKQDLAOCRLTCNDTSAALLISHIVQSGIGDF--D 172  
DB 121 FVYKFTPNPDLDEERYTRYLFTMQIKRLALGELCHSDNTASLLSAYLVQSCGDSSE 180  
QY 173 EALDREHLAKNKYIPOOD-ALEDKIVEFHHNHIGQTPAESDFOLLEIARLEMYGRLHP 231  
DB 181 DYPDATYLSHTFRVNPQTLFQFKKVMNDNRNFGMTPGESDLAMLEVARCOPFYGKLHA 240  
QY 232 AKDREGTKINLAVANTGILVFGQFTKINAFNNAKVRKLSFKRRFLIKLRPDANSAYQDT 291  
DB 241 AKDIDNGDAALSVHILGKIFVROLQDITFSWARIRKLSFKRRKLLVLPDPSQYLKET 300

```
QY 292 LEFTMASRDCCKSPWKICVHHAFRLFEPPKPKPKVLF-SRGSFRRSGRTQKQVLDY 350
Db 301 VEFSETRDECKNFKKCVCHHAFRCVQAEPPKPKKTRFFISKSSPRYHGRTKQKOLIDY 360
QY 351 VKEGGHKKVQFER 363
Db 361 VREHKKRREPFTTR 373

RESULT 5
O70318 PRELIMINARY; PRT; 988 AA.
AC O70318;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PROTEIN 4.1G (FRAGMENT).
GN EPB4.1L2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98198473; PubMed=9531554;
RA Walensky L.D., Gascard P., Fields M.E., Blackshaw S., Conboy J.G.,
RA Mohandas N., Snyder S.H.;
RT "The 13-kD FK506 binding protein, FKBP13, interacts with a novel
RT homologue of the erythrocyte membrane cytoskeletal protein 4.1.";
RL J. Cell Biol. 141:143-153(1998).
DR EMBL; AF044312; AAC40083.1;
DR MGD; MGI:103009; Epb4.1L2.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_4.1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00650; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00657; BAND_41_3; 1.
FT NON_TER 988
SQ SEQUENCE 988 AA; 109833 MW; B0367A16C5A2EC05 CRC64;

Query Match 35.8%; Score 699.5; DB 11; Length 988;
Best Local Similarity 43.7%; Pred. No. 1.1e-47;
Matches 150; Conservative 56; Mismatches 132; Indels 5; Gaps 5;

QY 34 TSPGKLVSIKIQMLDQTOEAFVQAPQAPGVLLDVCNHLNLVGDYFGLFPPDHKKITV 93
Db 205 TKTKTKVLAKVTLDDGTEYSCDLEKRAKGQVDFRCEHLNLEKDYFGLLFQDHPEQN 264
QY 94 WLDLKPIVKQIRRRKHVVVKKFPPDPDHTQQLBELTRYLFALQVQKODLAOGLTVCND 153
Db 265 WLDPKAEIKQLKNLWLFV-FNVKYPDPDSQLTEDITRYFLCLQRLQDIASGRPCSF 323
QY 154 TSAALLISHIVQSEIGDFD-EALDREHLAKNKYIP-QQDALEDKIVEFHNNHIGOTPAES 211
Db 324 VTALLGSYTLQAEHGDDPEYDSIDLQGFAPAHTRKELEKVSSELHKTTHRLSPAQA 383
QY 212 DFQLELTARLEMYGTLHPAKDREGTKINLAVANTGILVFGQFTKINAFNWKVRKLSF 271
Db 384 DSQFLENKRLSMYGVLDLHAKDSEGVDIKLGVCANGLLIYKDLRINRFAWPKILKISY 443
QY 272 KRRFLIKLRPDANSAYQDTLEFLMASRDCCKSPWKICVHHAFRLFEPPKPKPKVLF 331
Db 444 KRSNFIKVRPALEQEFESTIGFKLPNHRAAKRLKVKVCVEHHTFYRLV-SPEQPPKTKFL 502
QY 332 SRGSFRRSGRTQKQVLDYVKEGGHKKVQFERKHSIHRSLS 374
Db 503 TLGSKFRYSGRTQAOQREASTLIDRAPQOFERASSKRVS-RSL 544

RESULT 6
```

```
Q9WV92 PRELIMINARY; PRT; 929 AA.
AC Q9WV92;
DT 01-NOV-1999 (Tremblrel. 12, Created)
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PROTEIN 4.1B.
GN EPB4.1L3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN;
RA Parra M.K., Chan N.L., Gascard P.D., Walensky L.D., Lee G.L.,
RA Chasis J.A., Snyder S.H., Narla M., Conboy J.G.;
RT "Characterization of protein 4.1B, a new gene in the protein 4.1
RT family with high level, focal expression in brain.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF152247; AAD38048.1;
DR MGD; MGI:103008; Epb4.1L3.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00650; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00657; BAND_41_3; 1.
SQ SEQUENCE 929 AA; 103337 MW; F4975FF405DA44AE CRC64;

Query Match 35.4%; Score 691.5; DB 11; Length 929;
Best Local Similarity 41.9%; Pred. No. 4.5e-47;
Matches 153; Conservative 62; Mismatches 133; Indels 17; Gaps 9;

QY 18 ENSGISTLREQQ---KPPPTSGKLVSIKIQMLDQTOEAFVQAPQAPGVLLDVCNHL 73
Db 97 QRSSSSKLSRSLPKIVKRP-----KSMQCKVTLDDGSEYGVCDVKRSRGVLFDKVCEHL 151
QY 74 NLVEGDYFGLFPPDHKKITVWLDLKPVIKQIRRRK-HVVVKFVVFVFFPDHTQQLBELT 132
Db 152 NLEKDYFGLTYRDAENQKNWLDPAKEIKKQIRSGAWH--FSENVKFPYPPDPAQLSEDT 209
QY 133 RYLFALQVQKODLAOGLTVCNDTSAALLISHIVQSEIGDFD-EALDREHLAKNKYIPQ-Q 190
Db 210 RYLCQLQRLDDIVSGRLPCSFVTLALLGSYTVQSELGDYDPPDECNDYISERFANHTK 269
QY 191 ALEDKIVEFHNNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250
Db 270 ELEDKVIELHKSHRGMPAEAEHMFLENAKKLSMYGVLDLHAKDSEGVETMLGVCAAGLL 329
QY 251 VFQGTFTKINAFNWKVRKLSFKRKRFLIKLRPDANSAYQDTLEFLMASRDCCKSPWKICV 310
Db 330 IYDRRLRINRFAWPKVILKISYKRNNFYIKIRPGEFFQEFESTIGFKLPNHRAAKRLKVCV 389
QY 311 EHHAFRLFEPPKPKPKVLF-SRGSFRRSGRTQKQVLDYVKEGGHKKVQFERKHSIHRS 370
Db 390 EHHFTFRL-LPEAPPKFL-TLGSKFRYSGRTQAOQRRASALIDRAPYFFERSKRYT 447
QY 371 I-RSL 374
Db 448 MSRSL 452

RESULT 7
Q9JMB2 PRELIMINARY; PRT; 963 AA.
AC Q9JMB2;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE TYPE II BRAIN 4.1 MINOR ISOFORM.
```

OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;  
RX MEDLINE=20267856; PubMed=10806359;  
RA Yamakawa H., Ohara O.;  
RT "Comparison of mRNA and protein levels of four members of the protein  
RT 4.1 family: The type II brain 4.1/4.1b/KIAA0987 is the most  
RT predominant member of the protein 4.1 family in rat brain.";  
RL Gene 248:137-145(2000).  
DR EMBL; AB032828; BAA90775.1; -.  
DR InterPro; IPR000299; Band\_4.1.  
DR Pfam; PF00373; Band\_41; 1.  
DR PRINTS; PRO0935; BAND41.  
DR SMART; SM00295; B41; 1.  
DR PROSITE; PS00660; BAND\_41\_1; 1.  
DR PROSITE; PS00661; BAND\_41\_2; 1.  
DR PROSITE; PS50057; BAND\_41\_3; 1.  
SQ SEQUENCE 963 AA; 107071 MW; B37C137B04305F71 CRC64;  
  
Query Match 35.4%; Score 691.5; DB 11; Length 963;  
Best Local Similarity 41.9%; Pred. No. 4.8e-47;  
Matches 153; Conservative 62; Mismatches 133; Indels 17; Gaps 9;  
  
QY 18 ENSGISTLERGQ----KPPPTPSGKLVSITKIOMLDDTQEAFAFVQAPACKVLLDAVCNHL 73  
DB 92 QKSSSKLSRSLKIVKRP-----KSMQCKVTLDDGSEYGCDDVKRSRGQVLFDKVCEHL 146  
  
QY 74 NLVEGYFGLFPPDHKKITVWLDLKPVIKQIRRPK-HVVVKFVVFPPDPHTQLOEELT 132  
DB 147 NLEKDYFGLTYRDAENQKNWLDPAKEIKKQIRSGAWH--FSENVKVFPPDPAQLSEDI 204  
  
QY 133 RYLFALQVQDLAQRCLTCDNTSAALLSHIVQSEIGDGF-D-EALDREHLAKNKYIPQ-Q 190  
DB 205 RYLCQLRDDIDVSGRLPCSFVTLLGSGYVQSELGDDPDCEGNDYISEFRFAPNHTK 264  
  
QY 191 ALEDKIVEFHNNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250  
DB 265 ELEDKVIELKSHRGWTPAEAEHMFLENAKKLSMYGVDLHAKDSEGVIMLGVCASGLL 324  
  
QY 251 VFQGTFTKINAFNWKVLSFKRRKFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICV 310  
DB 325 IYRDLRLINRFAMPKVKLSYKRNIFYIKIRGCEFFQFESTIGFKLPNHRAAKRLWKVCV 384  
  
QY 311 EHHAFRLFEKPKPKPVLFSGRSFSGRTQKQVLDYVKEGGHKVKQVFERKHKSHS 370  
DB 385 EHHTFFRL-LPEAPPKKFL-TLGSFRYSGRTOAOTRRASALIDRAPAPYFERSSSKRYT 442  
  
QY 371 I-RSL 374  
DB 443 MSRL 447  
  
RESULT 8  
Q9JMB3  
ID Q9JMB3 PRELIMINARY; PRT: 1105 AA.  
AC Q9JMB3;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE TYPE II BRAIN 4.1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;  
RX MEDLINE=20267856; PubMed=10806359;

RA Yamakawa H., Ohara O.;  
RT "Comparison of mRNA and protein levels of four members of the protein  
RT 4.1 family: The type II brain 4.1/4.1b/KIAA0987 is the most  
RT predominant member of the protein 4.1 family in rat brain.";  
RL Gene 248:137-145(2000).  
DR EMBL; AB032827; BAA90774.1; -.  
DR InterPro; IPR000299; Band\_4.1.  
DR Pfam; PF00373; Band\_41; 1.  
DR PRINTS; PRO0935; BAND41.  
DR SMART; SM00295; B41; 1.  
DR PROSITE; PS00660; BAND\_41\_1; 1.  
DR PROSITE; PS00661; BAND\_41\_2; 1.  
DR PROSITE; PS50057; BAND\_41\_3; 1.  
SQ SEQUENCE 1105 AA; 122451 MW; 94FICA93A83693C6 CRC64;  
  
Query Match 35.4%; Score 691.5; DB 11; Length 1105;  
Best Local Similarity 41.9%; Pred. No. 5.7e-47;  
Matches 153; Conservative 62; Mismatches 133; Indels 17; Gaps 9;  
  
QY 18 ENSGISTLERGQ----KPPPTPSGKLVSITKIOMLDDTQEAFAFVQAPACKVLLDAVCNHL 73  
DB 92 QKSSSKLSRSLKIVKRP-----KSMQCKVTLDDGSEYGCDDVKRSRGQVLFDKVCEHL 146  
  
QY 74 NLVEGYFGLFPPDHKKITVWLDLKPVIKQIRRPK-HVVVKFVVFPPDPHTQLOEELT 132  
DB 147 NLEKDYFGLTYRDAENQKNWLDPAKEIKKQIRSGAWH--FSENVKVFPPDPAQLSEDI 204  
  
QY 133 RYLFALQVQDLAQRCLTCDNTSAALLSHIVQSEIGDGF-D-EALDREHLAKNKYIPQ-Q 190  
DB 205 RYLCQLRDDIDVSGRLPCSFVTLLGSGYVQSELGDDPDCEGNDYISEFRFAPNHTK 264  
  
QY 191 ALEDKIVEFHNNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250  
DB 265 ELEDKVIELKSHRGWTPAEAEHMFLENAKKLSMYGVDLHAKDSEGVIMLGVCASGLL 324  
  
QY 251 VFQGTFTKINAFNWKVLSFKRRKFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICV 310  
DB 325 IYRDLRLINRFAMPKVKLSYKRNIFYIKIRGCEFFQFESTIGFKLPNHRAAKRLWKVCV 384  
  
QY 311 EHHAFRLFEKPKPKPVLFSGRSFSGRTQKQVLDYVKEGGHKVKQVFERKHKSHS 370  
DB 385 EHHTFFRL-LPEAPPKKFL-TLGSFRYSGRTOAOTRRASALIDRAPAPYFERSSSKRYT 442  
  
QY 371 I-RSL 374  
DB 443 MSRL 447  
  
RESULT 9  
Q9BRP5  
ID Q9BRP5 PRELIMINARY; PRT: 865 AA.  
AC Q9BRP5;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
DE SIMILAR TO ERYTHROCYTE PROTEIN BAND 4.1-LIKE 3.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=LUNG CARCINOMA;  
RA Strausberg R.;  
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC006141; RAH06141.1; -.  
SQ SEQUENCE 865 AA; 96513 MW; C2070B01BF7F9422 CRC64;  
  
Query Match 35.3%; Score 690.5; DB 4; Length 865;  
Best Local Similarity 41.5%; Pred. No. 5e-47;  
Matches 152; Conservative 64; Mismatches 131; Indels 19; Gaps 9;

```
QY 18 ENSGISTLERG-----OKPPPTSGKLVSIKIQMLDDTQEAFAFVQAPAGKVLDDAVCNH 72
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 89 QKSSSKLSRSLPKIVKKP-----KSMQCKVILLDGSEYTCDEKRSRQGVLFDKVCEH 142

QY 73 LNLVEGDFGLPEPDHKKITVWLDLKPVIKQIRRPK-HVVVFVKVFFPPDHTQLQEEL 131
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 143 LNLLEKDYFGLTYRDAENQKNWLDPAKEIKKQVRSNAWH--FSFNKVFYPPDPAQLSEDI 200

QY 132 TRYLFALQVKDLOAQGLTCDNTSAALLISHIVQSEIGDFD-EALDREHLAKNKYIPQO- 189
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 229 TRYFLCQLRDDIVSGRLPCSFVTLALLGYSYTVQSELGDDYDPCGSDYISEFRFAPNHT 288

QY 190 DALEDKIVEFHNNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGI 249
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 201 TRYFLCQLRDDIVSGRLPCSFVTLALLGYSYTVQSELGDDYDPCGSDYISEFRFAPNHT 260

QY 190 DALEDKIVEFHNNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGI 249
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 261 KELEDKIVIELHKSHRGTPAEAEHMFLENAKKLSMVGVDLHAKDSEGVIMLGVCSAGL 320

QY 250 LVFGQFTKINAFNAWKVRLSKFRKRLIKLRPDANSAYQDTLEFLMASRDFCKSFWKIC 309
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 321 LIYDRRLIRNFPAWKPKVLFSGRSFSGRTQKQVLDYVKEGGHKKVQFERKSKYK 380

QY 310 VEHHAFERLEFEEPKPKPVLFSGRSFSGRTQKQVLDYVKEGGHKKVQFERKSKYK 369
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 381 VEHTTFERLL-LPEAPPKKFL-TLGSKFRYSRGTQAQTRRASALIDRPAPYFERSKRY 438

QY 370 SI-RSL 374
: | | | | |
Db 439 TMSRSL 444

RESULT 10
QY2J2 ID QY2J2 PRELIMINARY; PRT; 1115 AA.
AC QY2J2
DT 01-NOV-1999 (T-EMBLrel. 12, Created)
DT 01-NOV-1999 (T-EMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE KIAA0987 PROTEIN (FRAGMENT).
GN KIAA0987.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN.
RX MEDLINE=9246063; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.;"
RL DNA Res 6:63-70(1999).
DR EMBL; AB023204; BAA76831.1; -.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00657; BAND_41_3; 1.
DR SMART; SM00295; B41; 1.
FT NON_TER 1
SQ SEQUENCE 1115 AA; 123704 MW; E38E3A31E028D4A2 CRC64;
```

```
Query Match 35.3%; Score 690.5; DB 4; Length 1115;
Best Local Similarity 41.5%; Pred. No. 7e-47;
Matches 152; Conservative 64; Mismatches 131; Indels 19; Gaps 9;

QY 18 ENSGISTLERG-----OKPPPTSGKLVSIKIQMLDDTQEAFAFVQAPAGKVLDDAVCNH 72
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 117 QKSSSKLSRSLPKIVKKP-----KSMQCKVILLDGSEYTCDEKRSRQGVLFDKVCEH 170
```

```
QY 73 LNLVEGDFGLPEPDHKKITVWLDLKPVIKQIRRPK-HVVVFVKVFFPPDHTQLQEEL 131
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 171 LNLLEKDYFGLTYRDAENQKNWLDPAKEIKKQVRSNAWH--FSFNKVFYPPDPAQLSEDI 228

QY 132 TRYLFALQVKDLOAQGLTCDNTSAALLISHIVQSEIGDFD-EALDREHLAKNKYIPQO- 189
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 229 TRYFLCQLRDDIVSGRLPCSFVTLALLGYSYTVQSELGDDYDPCGSDYISEFRFAPNHT 288

QY 190 DALEDKIVEFHNNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGI 249
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 289 KELEDKIVIELHKSHRGTPAEAEHMFLENAKKLSMVGVDLHAKDSEGVIMLGVCSAGL 348

QY 250 LVFGQFTKINAFNAWKVRLSKFRKRLIKLRPDANSAYQDTLEFLMASRDFCKSFWKIC 309
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 349 LIYDRRLIRNFPAWKPKVLFSGRSFSGRTQKQVLDYVKEGGHKKVQFERKSKYK 408

QY 310 VEHHAFERLEFEEPKPKPVLFSGRSFSGRTQKQVLDYVKEGGHKKVQFERKSKYK 369
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 409 VEHTTFERLL-LPEAPPKKFL-TLGSKFRYSRGTQAQTRRASALIDRPAPYFERSKRY 466

QY 370 SI-RSL 374
: | | | | |
Db 467 TMSRSL 472

RESULT 11
QY2J2 ID QY2J2 PRELIMINARY; PRT; 1005 AA.
AC QY2J2
DT 01-JUN-1998 (T-EMBLrel. 06, Created)
DT 01-JUN-1998 (T-EMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (T-EMBLrel. 17, Last annotation update)
DE PROTEIN 4.1-G.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9826080; PubMed=9598318;
RA Parra M., Gascard P., Walensky L.D., Snyder S.H., Mohandas N.,
RA Conboy J.G.;
RT "Cloning and characterization of 4.1G (EPB41L2), a new member of the
RT skeletal protein 4.1 (EPB41) gene family.;"
RL Genomics 49:298-306(1998).
DR EMBL; AF027299; RAC16923.1; -.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR PROSITE; PS00660; BAND_41_1; UNKNOWN_1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00657; BAND_41_3; 1.
DR SMART; SM00295; B41; 1.
SQ SEQUENCE 1005 AA; 112587 MW; E86CB17488F6045F CRC64;
```

```
Query Match 35.3%; Score 689.5; DB 4; Length 1005;
Best Local Similarity 42.8%; Pred. No. 7.3e-47;
Matches 148; Conservative 58; Mismatches 129; Indels 11; Gaps 6;
```

```
QY 34 TPSSGLVSIKIQMLDDTQEAFAFVQAPAGKVLDDAVCNHNLVVEGDFGLPEPDHKKITV 93
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 212 TKTKTVQCKVTLLDGTSEYSCDLEKHAQGVLFKVCHEHLNLEKDYFGLLFQESPEQKN 271

QY 94 WDLDLKPIVKQIRRPKHVVVVFVKVFFPPDHTQLQEELTRYLFALQVKDLOAQGLTCDN 153
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 272 WDLPAKEIKRQLRNLPWLT-FNVKVFYPPDPSQLTDITRYFLCQLQRLQDIASGRPCSF 330

QY 154 TSAALLISHIVQSEIGDFDEALDREH---LAKNKYIP-QQDALEDKIVEFHNNHIGOTP 208
: | | | | | : | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 331 VTHALLGSTLQELQGLDYD---PEEHGSDLSLSEFQFAPQTRELEKVAELKTHKGLSP 387

QY 209 AESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIILVFOGFTKINAFNAWKVR 268
```

```
Db 388 AADSQFLNARLSMYGVDLHAKDSEGVKLGVCANGLLIYKDRLRINRFAWPKIK 447
Qy 269 LSFRRRLIKLRDPANSAYQDLEFLMASRDFCSFKWIKCVHEHFAFFRLPEEPKPKP 328
Db 448 ISYKRSNFYKVPAALEQFESTIGFKLPNHRRAKRLMKVCVEHHTFRLV-SPEOPPKA 506
Qy 329 VLFSGSSFRSGRTQKQVLDYKVGKGGHKKVQPERKHSKIHSIRSL 374
Db 507 KFTLGSFRYSGRTOAQTRQASTLIDRPAPHFERTSSKRV-SRL 551

RESULT 12
Q95713 ID O95713 PRELIMINARY; PRT; 503 AA.
AC O95713;
DT 01-MAY-1999 (TREMREL. 10, Created)
DT 01-MAY-1999 (TREMREL. 10, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE PUTATIVE LUNG TUMOR SUPPRESSOR.
GN DALI.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RX MEDLINE=991071198; PubMed=9892180;
RA Tran Y.K., Bogler O., Gorse K.M., Wieland I., Green M.R.,
RA Newsham I.F.;
RT "A novel member of the NF2/ERM/4.1 superfamily with growth suppressing
RT properties in lung cancer.";
RL Cancer Res. 59:35-43(1999).
DR EMBL: AF069072; AAC79806.1; -.
DR InterPro: IPR000299; Band_4.1.
DR Pfam: PF00373; Band_41; 1.
DR PRINTS: PR00935; BAND41.
DR PROSITE: PS00660; BAND_41_1; 1.
DR PROSITE: PS00661; BAND_41_2; 1.
DR PROSITE: PS00571; BAND_41_3; 1.
DR SMART: SM00295; B41; 1.
SQ SEQUENCE 503 AA; 57691 MW; 2B3E3F8142EA83AE CRC64;

Query Match 35.0%; Score 684; DB 4; Length 503;
Best Local Similarity 43.2%; Pred. No. 8e-47;
Matches 145; Conservative 61; Mismatches 122; Indels 8; Gaps 7;

Qy 43 KIOMLDDTQAEFVQAPGKVLDDAVCNHLNVEGDFGLFPPDHKKITVWDLKPIV 102
Db 4 KVLIDGSEYTCDEKRSRGVLFQKVCHELNLEKDYGLTVRDAENQKNWLDPAKEIK 63
Qy 103 KOIRPK-HVNVKVVVFPDHTQLOEELTRYLFALQVKQDLAQRLLCNDTSAALLIS 161
Db 64 KOVRGAWH--FSFNKVFPPDPAQLSEIDITRYLCLQRDDIVSGRFLTALLGS 121
Qy 162 HTVQSEIGDFD-EALDREHLANKYIPQO--DALEKIVFEHNNHIGOTPAESDFQLLETA 219
Db 122 YTVQSEIGDYDPDCEGSDYISERFAPNTKEKDVIELHKSHRGWTTPAEAMHLENA 181
Qy 220 RLEMYGIRLHPAKDREGTKINLAVANTGILVFGQGTIKINAFNWKVRLSKFRKRLIK 279
Db 182 KKLMSYGVDLHAKDSEGVLEMLGVASCGLLIYDRDLINRFAWPKVLSYKRNFFYIK 241
Qy 280 LRPDANSAYQDLEFLMASRDFCSFKWIKCVHEHFAFFRLPEEPKPKPVLSRGSFRF 339
Db 242 IRPEFEQFESTIGFKLPNHRRAKRLMKVCVEHHTFRLV-LPEAPPKFL-TLGSKFY 299
Qy 340 SGRTOQVLDYKVGKGGHKKVQPERKHSKIHSI-RSL 374
Db 300 SGRTOQTRASALIDRPAPYFERSSSKRYTWRSRL 335
```

```
RESULT 13
Q9Y579 ID Q9Y579 PRELIMINARY; PRT; 864 AA.
AC Q9Y579;
DT 01-NOV-1999 (TREMREL. 12, Created)
DT 01-NOV-1999 (TREMREL. 12, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE PROTEIN 4.1.
GN EPB4.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Huang S.C., Wang C., Lichtenauer U., Vortmeyer A., Zhuang Z.;
RT "Sequence of protein 4.1 from a human neuroblastoma cell line: LANS.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF156225; AAD42222.1; -.
DR InterPro: IPR000299; Band_4.1.
DR Pfam: PF00373; Band_41; 1.
DR PRINTS: PR00935; BAND41.
DR PROSITE: PS00660; BAND_41_1; 1.
DR PROSITE: PS00661; BAND_41_2; 1.
DR PROSITE: PS00571; BAND_41_3; 1.
DR SMART: SM00295; B41; 1.
SQ SEQUENCE 864 AA; 97056 MW; 49EDD4A686D97931 CRC64;

Query Match 34.8%; Score 681.5; DB 4; Length 864;
Best Local Similarity 42.0%; Pred. No. 2.6e-46;
Matches 152; Conservative 58; Mismatches 143; Indels 9; Gaps 7;

Qy 16 APENSGISTLERGQKPPPTPSGKLVSIKIQMLDDTQEAPEVQAPGKVLDDAVCNHLN 75
Db 188 SFQSKAETELKASQK--PIRKHRNMCKVSLDDTVECVVEKHAKQDQLLRVCEHLN 245
Qy 76 VEGDYFGLFPPDHKKITVWDLKPIVKQIR--PKHVNVKVVVFPDHTQLOEELTRY 134
Db 246 LEEDYFGLAIWNTTSTKWLDSAKEIKQVRGVPWNFT--FNVKFYPDPAQLTEDITRY 303
Qy 135 LFAQVKQDLAQRLLCNDTSAALLISHIVQSEIGDFDEALDR-EHLANKYIPQO-DAL 192
Db 304 YLCQLQRQDIVAGRLPCSFATLALLGSYTIQSELGVDYDPHGLGVYVSDFKLAPNQTKE 363
Qy 193 EKIIVEFHNNHIGOTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVF 252
Db 364 EKVVELHKSYSRMTQAQDLFELENAKKLSMTGVDLHRAKDLLEGVDIILGVCSGLLY 423
Qy 253 QGFTKINAFNWKVRLSKFRRLIKLRPDANSAYQDLEFLMASRDFCSFKWIKCVHEH 312
Db 424 KDKLRINRFPWPKVLSYKRSFFIKIRPGEQEQESTIGFKLPVSRRAKKLWKVCVEH 483
Qy 313 HAFFRLFEPEKPKPVLSRGSFRSGRTQKQVLDYKVGKGGHKKVQPERKHSKIHSIR 372
Db 484 HTFFRL-TSTDITPKSKFALGSKFRYSGRTOAQTRQASALIDRPAPHFERTASK-RASR 541
Qy 373 SL 374
Db 542 SL 543

RESULT 14
Q9R102 ID Q9R102 PRELIMINARY; PRT; 706 AA.
AC Q9R102;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-JUN-2001 (TREMREL. 17, Last annotation update)
DE DALI.
GN DALI.
OS Mus musculus (Mouse).
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:50:18 ; Search time 36.72 Seconds  
(without alignments)  
640.412 Million cell updates/sec

Title: US-09-555-342A-2  
Perfect score: 5463  
Sequence: 1 MGIEQRPTGSLGAPENS.....SATSSASRPHVLKSHESLVY 1045

Scoring table:  
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	646	11.8	345	2	US-08-446-345-40
2	612	11.2	288	3	US-09-100-804-14
3	522	9.6	247	3	US-09-100-804-13
4	500.5	9.2	288	3	US-09-100-804-15
5	499	9.1	342	2	US-08-446-345-39
6	396	7.2	1174	2	US-08-446-345-36
7	344	6.3	296	3	US-09-100-804-12
8	335	6.1	591	1	US-08-179-738-10
9	335	6.1	591	2	US-08-628-145-10
10	335	6.1	595	1	US-08-171-718-16
11	335	6.1	595	3	US-08-478-087-16
12	335	6.1	596	1	US-08-179-738-3
13	335	6.1	596	2	US-08-628-145-3
14	332	6.1	584	1	US-08-179-738-7
15	332	6.1	584	2	US-08-628-145-7
16	330	6.0	591	1	US-08-179-738-5
17	330	6.0	591	2	US-08-628-145-5
18	330	6.0	596	1	US-08-179-738-2
19	330	6.0	596	2	US-08-628-145-2
20	308	5.6	2485	4	US-09-290-640-46
21	300.5	5.5	2466	3	US-09-080-855-12
22	300.5	5.5	2466	5	PCT-US94-09943-2
23	296	5.4	2465	2	US-08-596-291-3
24	296	5.4	2465	3	US-09-100-804-3
25	265	4.9	303	3	US-09-100-804-11
26	264.5	4.8	2860	2	US-08-826-267-2
27	212.5	3.9	1461	2	US-08-993-228-10

28	206	3.8	1244	3	US-09-356-952-7	Sequence 7, Appli
29	180	3.3	761	4	US-09-625-188-14	Sequence 14, Appli
30	179.5	3.3	912	4	US-08-943-768-2	Sequence 2, Appli
31	159.5	2.9	477	1	US-08-191-338A-2	Sequence 2, Appli
32	142	2.6	1319	2	US-08-290-731C-2	Sequence 2, Appli
33	142	2.6	1336	2	US-08-290-731C-6	Sequence 6, Appli
34	140	2.6	1297	2	US-08-290-731C-4	Sequence 4, Appli
35	137.5	2.5	2409	6	5180808-2	Patent No. 5180808
36	136	2.5	1333	3	US-09-356-952-2	Sequence 2, Appli
37	136	2.5	1805	1	US-07-853-913-2	Sequence 2, Appli
38	135.5	2.5	750	4	US-08-202-841A-2	Sequence 2, Appli
39	133.5	2.4	1354	3	US-08-685-871-2	Sequence 2, Appli
40	133	2.4	620	1	US-08-484-105-4	Sequence 4, Appli
41	133	2.4	620	1	US-08-484-106-4	Sequence 4, Appli
42	128	2.3	1572	2	US-08-290-731C-5	Sequence 5, Appli
43	128	2.3	1596	3	US-09-356-952-3	Sequence 3, Appli
44	126	2.3	772	1	US-08-524-757-12	Sequence 12, Appli
45	122.5	2.2	399	4	US-09-284-033-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-446-345-40  
; Sequence 40, Application US/08446345  
; Patent No. 5831009  
; GENERAL INFORMATION:  
; APPLICANT: Ullrich, Axel  
; APPLICANT: Moller, Niels P.H.  
; APPLICANT: Moller, Karin B.  
; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: N.Y.  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,345  
; FILING DATE: 22-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/234,440  
; FILING DATE: 28-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30742  
; REFERENCE/DOCKET NUMBER: 7683-054  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 40:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 345 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: unknown  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
US-08-446-345-40

Query Match 11.8%; Score 646; DB 2; Length 345;  
Best Local Similarity 40.9%; Pred. No. 9e-47;

Matches 139; Conservative 62; Mismatches 115; Indels 24; Gaps 7;

QY 35 PSGLYSIK-----IQLLDDTQEAPEVQAPRGKVLDDAVCNHNLVVG 78  
 Db 8 PACTYVRASELARBOHTTEVCNILLDNTVQAEKVNKHQGVLLDVVFKHLDLQTEQ 67  
 QY 79 DYGLFPPDHKKITV-WLDLLKPIVKQIRPKHVHVVKVFFPPDHTQLQELTRYLFA 137  
 Db 68 DYGLQLADSDTNPRLDNPKNPIRQLKRGSPYSLNFRVFPVDPNKLQEEYTRYQF 127  
 QY 138 LQVKQDLAQRGRLTCDNTSAALLSHIVQSEIGDFDEALDRE-HLAKNKYIPQO-DALEBK 195  
 Db 128 LQIKQDILTRGLPCPNTAALL-AFAVQSELGDDYQSENLSGLSYDFIPNQDFEKE 186  
 QY 196 IVFHHNHIGOTPAESDFQLLEATRLMVGIRLHPAKDREGTKINLAVANTGILVFOGF 255  
 Db 187 IAKLHQHIGLSPAEAEFNVNTARTLELYGVFEHYARDOSNEIMGVMSGGLIYKNR 246  
 QY 256 TKINAFNWKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHAF 315  
 Db 247 VRNTEPWLKIVKISFKCKOFFQLRKELHESRETLGFMNMYRACKNLWKACVEHHTF 306  
 QY 316 FRLPEPKPKPVL---PSRGSFRESGRTQKQVLDYVK 352  
 Db 307 FRL-DRPLPPQKNFFAHYFTLGSKFRYCGRTEVQSVQYK 345

RESULT 2

US-09-100-804-14  
 ; Sequence 14, Application US/09100804  
 ; Patent No. 6066472  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GONEZ, LEONEL JORGE  
 ; APPLICANT: SARAS, JAN  
 ; APPLICANT: CLAESSON-WELSH, LENA  
 ; APPLICANT: HELDIN, CARL-HENRIK  
 ; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
 ; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN  
 ; TITLE OF INVENTION: TYROSINE PHOSPHATASES  
 ; NUMBER OF SEQUENCES: 34  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
 ; STREET: 600 ATLANTIC AVENUE  
 ; CITY: BOSTON  
 ; STATE: MASSACHUSETTS  
 ; COUNTRY: USA  
 ; ZIP: 02210  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/100,804  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/596,291  
 ; FILING DATE: 09-AUG-1996  
 ; APPLICATION NUMBER: US 08/115,573  
 ; FILING DATE: 01-SEP-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US94/09943  
 ; FILING DATE: 01-SEP-1994  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: GATES, EDWARD R.  
 ; REGISTRATION NUMBER: 31,616  
 ; REFERENCE/DOCKET NUMBER: LQ461/7003  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 617-720-3500  
 ; TELEFAX: 617-720-2441  
 ; TELEX:  
 ; INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 288 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 US-09-100-804-14

Query Match 11.2%; Score 612; DB 3; Length 288;  
 Best Local Similarity 44.1%; Pred. No. 5.2e-44;  
 Matches 127; Conservative 52; Mismatches 105; Indels 4; Gaps 4;  
 QY 40 VSIQIQLDDTQEAPEVQAPRGKVLDDAVCNHNLVVGDFGLFPPDHKKITV-WLDLL 98  
 Db 1 VVCNILLDNTVQAEKVNKHQGVLLDVVFKHLDLQTEQDYFGLQLADSDTNPRLDNP 60  
 QY 99 KPIVKQIRPKHVHVVKVFFPPDHTQLQELTRYLFAQLQVKQDLAQRGRLTCDNTSAAL 158  
 Db 61 KPIRQLKRGSPYSLNFRVFPVDPNKLQEEYTRYQYFLQIKQDILTRGLPCPSNTAAL 120  
 QY 159 LISHIVQSEIGDFDEALDRE-HLAKNKYIPQO-DALEDKIVEPHHNIQOTPAESDFQL 216  
 Db 121 LASFAVQSELGDDYQSENLSGLSYDFIPNQDFEKEIAKLHQHIGLSPAEAEFNV 180  
 QY 217 ETARLEMYGIRLHPAKDREGTKINLAVANTGILVFOGFTKINAFNWKVRKLSFKRRF 276  
 Db 181 NTARTLELYGVFEHYARDOSNEIMGVMSGGLIYKNRVMNTFPMLKIVKISFKCKOF 240  
 QY 277 LIKLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHAFRLPEPKP 324  
 Db 241 FQLRKELHESRETLGFMNMYRACKNLWKACVEHHTFRL-DRPLP 287

RESULT 3

US-09-100-804-13  
 ; Sequence 13, Application US/09100804  
 ; Patent No. 6066472  
 ; GENERAL INFORMATION:  
 ; APPLICANT: GONEZ, LEONEL JORGE  
 ; APPLICANT: SARAS, JAN  
 ; APPLICANT: CLAESSON-WELSH, LENA  
 ; APPLICANT: HELDIN, CARL-HENRIK  
 ; TITLE OF INVENTION: PRIMARY STRUCTURE AND FUNCTIONAL  
 ; TITLE OF INVENTION: EXPRESSION OF NUCLEOTIDE SEQUENCES FOR NOVEL PROTEIN  
 ; TITLE OF INVENTION: TYROSINE PHOSPHATASES  
 ; NUMBER OF SEQUENCES: 34  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: WOLF, GREENFIELD & SACKS, P.C.  
 ; STREET: 600 ATLANTIC AVENUE  
 ; CITY: BOSTON  
 ; STATE: MASSACHUSETTS  
 ; COUNTRY: USA  
 ; ZIP: 02210  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/100,804  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/596,291  
 ; FILING DATE: 09-AUG-1996  
 ; APPLICATION NUMBER: US 08/115,573  
 ; FILING DATE: 01-SEP-1993  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US94/09943  
 ; FILING DATE: 01-SEP-1994



[illegible]

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446, 345  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/234, 440  
FILING DATE: 28-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30742  
REFERENCE/DOCKET NUMBER: 7683-054  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 342 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-446-345-39

Query Match 9.1%; Score 499; DB 2; Length 342;  
Best Local Similarity 34.7%; Pred. No. 2.9e-34;  
Matches 114; Conservative 66; Mismatches 133; Indels 16; Gaps 7;

QY 27 RGQPPPTPSGKLVSIKIQMLDDTQEAPEVPQAPGKVLDDAVCNHNLNVEGDFGLFP 86  
Db 16 RTSLEPKETSEVICSIFHLDGQVQFVKTKQDTGGVLLDMVHNLGVTEKEIFGLQHD 75  
QY 87 DHKKITV-WLDLLKPIVKQIRRPKHVVVVKVFFPDHPTQLQBELTRYLFALQVKQDLA 145  
Db 76 DSDVSPRWLEASLPKQLKGGFPCTLHFRVRFIPDPNTLQEQTRHLYFLQLKMDIC 135  
QY 146 QGRUTCNTSNAALLSHIVQSEIGDFDEALDRE-HLAKNKYIQQDALED---KIVEFHH 201  
Db 136 EGRUTCPLNVAVLASVAVOSGHFDYNSHTHPGYLSDSHFIPDN--EDFLTKVESLHE 193  
QY 202 NHIGQTPAESDFQLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFGFTKINAF 261  
Db 194 QHSLKQSEAEVCYINTARTLDFYGVHLSGRDLHNLDMIGASAGVAYRYKICTSFY 253  
QY 262 NNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSWKTCVHEHFAFFRLFE 321  
Db 254 PWNILKISPKRKKFFTHQKQAESREHIVAFNMLNRYRCKNLWKSCVBEHHTFFQA-KK 312  
QY 322 PKPKPKVL---PSRGSSFRSGRTQKV 347  
Db 313 LLFQEKNVLSQYWTMGR-----NTRKSV 336

RESULT 6  
US-08-446-345-36  
Sequence 36, Application US/08446345  
Patent No. 5831009  
GENERAL INFORMATION:  
APPLICANT: Ullrich, Axel  
APPLICANT: Moller, Niels P.H.  
APPLICANT: Moller, Karin B.  
TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE  
TITLE OF INVENTION: PHOSPHATASES PTP-DI  
NUMBER OF SEQUENCES: 41  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York

STATE: N.Y.  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent in Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/446, 345  
FILING DATE: 22-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/234, 440  
FILING DATE: 28-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Coruzzi, Laura A.  
REGISTRATION NUMBER: 30742  
REFERENCE/DOCKET NUMBER: 7683-054  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1174 amino acids  
TYPE: amino acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-446-345-36

Query Match 7.2%; Score 396; DB 2; Length 1174;  
Best Local Similarity 22.5%; Pred. No. 1.3e-24;  
Matches 141; Conservative 103; Mismatches 248; Indels 134; Gaps 18;

QY 25 LERQKPPPTPSGKLVSIKIQMLDDTQEAPEVPQAPGKVLDDAVCNHNLNVEGDFGLFE 84  
Db 9 LKTRTYVSSKCLVA-RIOLLNNEFEETLSVESTGQESLEAVAQRLELREVTFSLW 67  
QY 85 FPDHKITVWLDLLKPIVKQIRRPKHVVVVKVFFPDHPTQLQBELTRYLFALQVKQD 143  
Db 68 YNKNQQRWVLDLEKPKQLDKYALEPTVYFGVWFVPSVLSQLOQEIYRYQLQKLD 127  
QY 144 LAQGRUTCNTSNAALLSHIVQSEIGDFDEALDREHLAKNKYIP---QQDALED---K 195  
Db 128 ILEGSIPCTLEQAIQIAGLAVQADFGDFQYESQDFLQRFALFPVGLQDEKVKLEATQK 187  
QY 196 IVEFHHNHIGQTPAESDFQLEIARLEMYGIRLHPAKDREGTKINLAVANTGILV-FQG 254  
Db 188 VALLHQYKRGLTAPDAEMLYMQVERMDGYGESYPAKDSQSGDISIGACLEGIFVKHN 247  
QY 255 FYKINAFNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSWKTCVHEHHA 314  
Db 248 GRHPVVFRWHDIANMGNHNSFFALEL---ANK--EETIQFQTEDEMTAKYIWRCLVARHK 302  
QY 315 FFRLEP-----EPKPKPKVLFSRGSSFRPSGR-----TQKVL 348  
Db 303 FYRLNOCNLTQTVTVNPTRRSSRMSLPKQPYVYVMPPPPQLHYNGHYTEPYASSQDNL 362  
QY 349 DYVKEGG---HKVQOFER-----KHSKHSIKS-----LASOPTEL 381  
Db 363 FVPNQNGYCHSOTSILDRADIFNGRIRNGSVYSAHSTNSLNNPQYLPQSPMSSNPSIT 422  
QY 382 NSEVL-----EESQOSTSLTFG-----398  
Db 423 GSDVMRPDYLPSHRHSAVIPPYRPTPDYETVMKQLNRLGLVHAERQSHSLRNLNIGSSYA 482  
QY 399 -----EGAESPGQSCRRGKEPKVSAGEPSPAPRSPAGNKQADGA 443  
Db 483 YSRPAALVYSQPEIREHAQLPSPAAAHCFSLSYSPHSPSPYPAERRPVGVAVSPEL 542



```
Db 12 SSLKR--KOPKTTFTVRIVTM-----DAEMFNCEMKWKGLDFDLVLCRTLGLRETWFFG 63
QY 83 LEPPDHKKITVWLDLKPPIV-KOIRRPKHVVVVKVFFPPD-HTLOEELTRYLFALOV 140
Db 64 LQY-TIKDVANLKMCKKVLHDVSKKEEPTVFHFLAKFYFENAEELVQBITQHLFFLOV 122
QY 141 KDLAQRRLTCNDTSAAALLSHIVQSEIGDFDEAL-DREHLAKNKYIPQO-----D 190
Db 123 KKOILDEKIVCPPEASVLLASVAVQAKYGDYDPSVHKRGFLAQOELLPKRVINLYQMTPE 182
QY 191 ALEDKIVFHHNIGQTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250
Db 183 MWEERTAMTAERGRARDEAEYKIAQDLEMYGVNYFAIRNKGTGELLGVDAALGLH 242
QY 251 VFOGFTKIN--AFNNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWK 307
Db 243 IYDPENRLTPKISFPWNEIRNISYSDEFTIK-----PLDKKIDVFNSSKLRVNLKILQ 298
QY 308 ICVEHHAFFRLFEERPKPKPVLFSGSSFRFGRTQKQVLDVYVKEGHHKKVQFERKHSK 367
Db 299 LCIGNHDLF--MRRRKADSLEVOOMKAQAREEKARQMERQRLARE---KOMREEAERTR 353
QY 368 IHSIRSLASOPT--LNSEVLEQSQSTSLTFEGEASPGQSCRRCKEKPVSAGEPGSH 425
Db 354 DELERRLLQKKEATMANEALMRSEETADL-LAEKAQI-----TEEEAKLLAQAAEA 405
QY 426 PSPAPRRSPAGNKQADGAASAPTEEEVVKVDRTOOSK 463
Db 406 EQEMQRIK-----ATAIRTEEEKRLMEQKVLAE 434

RESULT 9
US-08-628-145-10
; Sequence 10, Application US/08628145
; Patent No. 5872214
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kley, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5872214el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,145
; FILING DATE: 04-APR-1996
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,738
; FILING DATE: 10-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 591 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

```
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: Homo sapien
; US-08-628-145-10

Query Match 6.1%; Score 335; DB 2; Length 591;
Best Local Similarity 23.8%; Pred. No. 6.3e-20;
Matches 109; Conservative 106; Mismatches 191; Indels 52; Gaps 15;

QY 23 STLERQKPPPTSGKLIVSIKIQMLDDTQFAFEVQAPQAPKVLDDAVCNHLNLVEGDYFG 82
Db 12 SSLKR--KOPKTTFTVRIVTM-----DAEMFNCEMKWKGLDFDLVLCRTLGLRETWFFG 63
QY 83 LEPPDHKKITVWLDLKPPIV-KOIRRPKHVVVVKVFFPPD-HTLOEELTRYLFALOV 140
Db 64 LQY-TIKDVANLKMCKKVLHDVSKKEEPTVFHFLAKFYFENAEELVQBITQHLFFLOV 122
QY 141 KDLAQRRLTCNDTSAAALLSHIVQSEIGDFDEAL-DREHLAKNKYIPQO-----D 190
Db 123 KKOILDEKIVCPPEASVLLASVAVQAKYGDYDPSVHKRGFLAQOELLPKRVINLYQMTPE 182
QY 191 ALEDKIVFHHNIGQTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGIL 250
Db 183 MWEERTAMTAERGRARDEAEYKIAQDLEMYGVNYFAIRNKGTGELLGVDAALGLH 242
QY 251 VFOGFTKIN--AFNNAKVRKLSFKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFWK 307
Db 243 IYDPENRLTPKISFPWNEIRNISYSDEFTIK-----PLDKKIDVFNSSKLRVNLKILQ 298
QY 308 ICVEHHAFFRLFEERPKPKPVLFSGSSFRFGRTQKQVLDVYVKEGHHKKVQFERKHSK 367
Db 299 LCIGNHDLF--MRRRKADSLEVOOMKAQAREEKARQMERQRLARE---KOMREEAERTR 353
QY 368 IHSIRSLASOPT--LNSEVLEQSQSTSLTFEGEASPGQSCRRCKEKPVSAGEPGSH 425
Db 354 DELERRLLQKKEATMANEALMRSEETADL-LAEKAQI-----TEEEAKLLAQAAEA 405
QY 426 PSPAPRRSPAGNKQADGAASAPTEEEVVKVDRTOOSK 463
Db 406 EQEMQRIK-----ATAIRTEEEKRLMEQKVLAE 434

RESULT 10
US-08-171-718-16
; Sequence 16, Application US/08171718
; Patent No. 5707863
; GENERAL INFORMATION:
; APPLICANT: Trofatter, James A.
; APPLICANT: MacCollin, Mia M.
; APPLICANT: Gusella, James F.
; TITLE OF INVENTION: Tumor Suppressor Gene Merlin and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/171,718
; FILING DATE: 22-DEC-1993
; CLASSIFICATION: 436
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/108,808
; FILING DATE: 19-AUG-1993
```

```
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/022,034
; FILING DATE: 25-FEB-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/026,063
; FILING DATE: 04-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Anne
; REGISTRATION NUMBER: 36,463
; REFERENCE/DOCKET NUMBER: 0609.3850003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-171-718-16

Query Match 6.1%; Score 335; DB 1; Length 595;
Best Local Similarity 23.8%; Pred. No. 6.4e-20;
Matches 109; Conservative 106; Mismatches 191; Indels 52; Gaps 15;

QY 23 STLERGQKPPPTSGKLVSIKIQMLDDTQAEFVQAPGKVLDDAVCNHNLVGDYFG 82
DB 12 SSLKR--KQPKFTTVRIVTM-----DAEMFNCMKWKGKDLFDLVCRTLGLRETWFG 63
QY 83 LEFPDHKKITVWDLKPIV-KQIRPKHVVKVVFPPPD-HTQLQBELTRYLFALQV 140
DB 64 LQY-TIKDTVAVLKMCKVLDHVSKEEPTVTFHLAKFYPENAEELVQEIOTHLFFLQV 122
QY 141 KQDLAQRGLTCNDTSAALLISHIVQSEIGDFDEAL-DREHLAKNKVYPOQ-----D 190
DB 123 KQILDDEKICYPPEASVLLASVAVQAKYGDYDPSVHKRGFLAQEELLPRKRVINLYQMTPE 182
QY 191 ALEDKIVEPHHNHIGOTPAESDFQLEIARLEMYGIRLHPAKDREGTKINLAVANTGIL 250
DB 183 MWEERTAWYAEHGRGARDEAEWEYLKIAQDLEMVGVNFAIRNKKGTLLGLVDALGLH 242
QY 251 VFOGFTKIN---AFNWKVKLSFKRRLIKLRPDANSAYQDTLEFLMASRDFCKSFVK 307
DB 243 IYDPENRLTPKISFPWNEIRNISYSKTEFTIK-----PLDKKIDVFKFNSKLRVKNLIQ 298
QY 308 ICVEHHAFFRLFEPEKPKPVLFSGSSFRSGTQKQVLDYVKGEGHKKVQFERKHSK 367
DB 299 LCIGNHDLF--MRRRKADSLVQOMKAQAREKARKOMERQRLARE---KOMREAEARTR 353
QY 368 IHSIRSLASQPT--LNSEVLQSQOSTSLTFEGEASPGQSCRRGKPKVSAGEPGSH 425
DB 354 DELERRLLQMKKEATMANEALRSEETADL-LAEKRAQI-----TEEBAKLLAQKAAEA 405
QY 426 PSPAPRRSPAGNKQADGAASAPTEEEVVVKORTQOSK 463
DB 406 EOEMORIK-----ATAIRTEEEKRLMEQKVLAE 434

RESULT 11
US-08-478-087-16
; Sequence 16, Application US/08478087
; Patent No. 6077685
; GENERAL INFORMATION:
; APPLICANT: Trofater, James A.
; APPLICANT: MacCollin, Mia M.
; TITLE OF INVENTION: Tumor Suppressor Gene Merlin and Uses
; TITLE OF INVENTION: Thereof
; NUMBER OF SEQUENCES: 120
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, N.W., Suite 600
```

```
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,087
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/171,718
; FILING DATE: 22-DEC-1993
; APPLICATION NUMBER: US 08/108,808
; FILING DATE: 19-AUG-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/022,034
; FILING DATE: 25-FEB-1993
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/026,063
; ATTORNEY/AGENT INFORMATION:
; NAME: Brown, Anne
; REGISTRATION NUMBER: 36,463
; REFERENCE/DOCKET NUMBER: 0609.3850003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 595 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-087-16

Query Match 6.1%; Score 335; DB 3; Length 595;
Best Local Similarity 23.8%; Pred. No. 6.4e-20;
Matches 109; Conservative 106; Mismatches 191; Indels 52; Gaps 15;

QY 23 STLERGQKPPPTSGKLVSIKIQMLDDTQAEFVQAPGKVLDDAVCNHNLVGDYFG 82
DB 12 SSLKR--KQPKFTTVRIVTM-----DAEMFNCMKWKGKDLFDLVCRTLGLRETWFG 63
QY 83 LEFPDHKKITVWDLKPIV-KQIRPKHVVKVVFPPPD-HTQLQBELTRYLFALQV 140
DB 64 LQY-TIKDTVAVLKMCKVLDHVSKEEPTVTFHLAKFYPENAEELVQEIOTHLFFLQV 122
QY 141 KQDLAQRGLTCNDTSAALLISHIVQSEIGDFDEAL-DREHLAKNKVYPOQ-----D 190
DB 123 KQILDDEKICYPPEASVLLASVAVQAKYGDYDPSVHKRGFLAQEELLPRKRVINLYQMTPE 182
QY 191 ALEDKIVEPHHNHIGOTPAESDFQLEIARLEMYGIRLHPAKDREGTKINLAVANTGIL 250
DB 183 MWEERTAWYAEHGRGARDEAEWEYLKIAQDLEMVGVNFAIRNKKGTLLGLVDALGLH 242
QY 251 VFOGFTKIN---AFNWKVKLSFKRRLIKLRPDANSAYQDTLEFLMASRDFCKSFVK 307
DB 243 IYDPENRLTPKISFPWNEIRNISYSKTEFTIK-----PLDKKIDVFKFNSKLRVKNLIQ 298
QY 308 ICVEHHAFFRLFEPEKPKPVLFSGSSFRSGTQKQVLDYVKGEGHKKVQFERKHSK 367
DB 299 LCIGNHDLF--MRRRKADSLVQOMKAQAREKARKOMERQRLARE---KOMREAEARTR 353
QY 368 IHSIRSLASQPT--LNSEVLQSQOSTSLTFEGEASPGQSCRRGKPKVSAGEPGSH 425
DB 354 DELERRLLQMKKEATMANEALRSEETADL-LAEKRAQI-----TEEBAKLLAQKAAEA 405
QY 426 PSPAPRRSPAGNKQADGAASAPTEEEVVVKORTQOSK 463
```

Db 406 EQEMQRIK-----ATAIRTEEEKRLMEQKVLAE 434

## RESULT 12

US-08-179-738-3  
; Sequence 3, Application US/08179738  
; Patent No. 5578462  
; GENERAL INFORMATION:  
; APPLICANT: Seizinger, Bernd R.  
; APPLICANT: Kley, Nikolai A.  
; APPLICANT: Bianchi, Albert B.  
; TITLE OF INVENTION: No. 5578462el NF2 Isoforms  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Reed & Robins  
; STREET: 635 Bryant Street  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: U.S.A  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/179,738  
; FILING DATE: 10-JAN-1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Robins, Roberta L.  
; REGISTRATION NUMBER: 33,208  
; REFERENCE/DOCKET NUMBER: 5998-0017  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 617-8999  
; TELEFAX: (415) 327-3231  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 596 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapien  
; US-08-179-738-3

Query Match 6.1%; Score 335; DB 1; Length 596;  
Best Local Similarity 23.8%; Pred. No. 6.4e-20;  
Matches 109; Conservative 106; Mismatches 191; Indels 52; Gaps 15;

QY 23 STLERGQPPPTSGKLVSIKIQMLDDTQAFVQAPGKVLDDAVCNHNLNVEGDYFG 82  
Db 12 SSLKR--KQPKTFTVRITM-----DAEMEFNCMKWKGLFDLCVRLTGLRETWFFG 63  
QY 83 LEFPDHHKTIIVLDDLLKPIV-KQIRPKHVHVVFVVFPPD-HTQLOBELTYLALQV 140  
Db 64 LQY-TIKDTVAVLKMDDKVLDDHVDVSKPEPTVTFHFAKFPENAEELVQEIOTHLFFLQV 122  
QY 141 KDLAQRGLTCDNTSAALILSHIVQSEIGDFDEAL-DREHLAKNKVIPPQ-----D 190  
Db 123 KQILDEKIVCPPEASVLLASVYQAKYGDYDPSVHKRGLFAQELLPRKVINLYQMTPE 182  
QY 191 ALEDKIVFHHNHIGTPAESDFQLEIARRLEMYGIRLHPAKDRGTKINLAVANTGIL 250  
Db 183 MWERITAWTAERGRARDEAEYLLKIAQDLEMYGVNFYAIRNKKGTTELLLGVDALGLH 242  
QY 251 VEOGFTKIN--AFNNAKVRKLSFKRRLIKLRPDANSAYQDTLEFLMASRDFCSFWK 307  
Db 243 IYDPENRLTPKISFPNWEINRISYSKFEFTIK-----PLDKKIDVFKFNSSKLRVKNLILQ 298

QY 308 ICVEHHAFFRLFEPPKPKPVLFSGSGFRFSRGTQKQVLDVYVKEGGHKVQFERKHSK 367  
Db 299 LCIGNHDLF--MRRRKADSLVQOMKAQAREEKARQMERQRLARE---KOMREEAERTR 353  
QY 368 IHSIRSLASQPTB--LNSEVLEQSQSTSLTFGEAGAESPGGQSCRRGKPKVKSAGEPGSH 425  
Db 354 DELERLLQMKKEATMANEALMRSEETADL-LAEKQAI-----TEEAKLLLAQAAEA 405  
QY 426 PSPAPRRSPAGNKQADGAASAPTEEEVEVVKDRTQOSK 463  
Db 406 EQEMQRIK-----ATAIRTEEEKRLMEQKVLAE 434

## RESULT 13

US-08-628-145-3  
; Sequence 3, Application US/08628145  
; Patent No. 5872214  
; GENERAL INFORMATION:  
; APPLICANT: Seizinger, Bernd R.  
; APPLICANT: Kley, Nikolai A.  
; APPLICANT: Bianchi, Albert B.  
; TITLE OF INVENTION: No. 5872214el NF2 Isoforms  
; NUMBER OF SEQUENCES: 26  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Reed & Robins  
; STREET: 635 Bryant Street  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: U.S.A  
; ZIP: 94301  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/628,145  
; FILING DATE: 04-APR-1996  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/179,738  
; FILING DATE: 10-JAN-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Robins, Roberta L.  
; REGISTRATION NUMBER: 33,208  
; REFERENCE/DOCKET NUMBER: 5998-0017  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 617-8999  
; TELEFAX: (415) 327-3231  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 596 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapien  
; US-08-628-145-3

Query Match 6.1%; Score 335; DB 2; Length 596;

Best Local Similarity 23.8%; Pred. No. 6.4e-20;  
Matches 109; Conservative 106; Mismatches 191; Indels 52; Gaps 15;

QY 23 STLERGQPPPTSGKLVSIKIQMLDDTQAFVQAPGKVLDDAVCNHNLNVEGDYFG 82  
Db 12 SSLKR--KQPKTFTVRITM-----DAEMEFNCMKWKGLFDLCVRLTGLRETWFFG 63  
QY 83 LEFPDHHKTIIVLDDLLKPIV-KQIRPKHVHVVFVVFPPD-HTQLOBELTYLALQV 140  
Db 64 LQY-TIKDTVAVLKMDDKVLDDHVDVSKPEPTVTFHFAKFPENAEELVQEIOTHLFFLQV 122

```

RESULT 15
US-08-628-145-7
; Sequence 7, Application US/08628145
; Patent No. 5872214
; GENERAL INFORMATION:
; APPLICANT: Seizinger, Bernd R.
; APPLICANT: Kley, Nikolai A.
; APPLICANT: Bianchi, Albert B.
; TITLE OF INVENTION: No. 5872214el NF2 Isoforms
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Reed & Robins
; STREET: 635 Bryant Street
; CITY: Palo Alto
; STATE: California
; COUNTRY: U.S.A
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,145
; FILING DATE: 04-APR-1996
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/179,738
; FILING DATE: 10-JAN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Robins, Roberta L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 584 amino acids

```

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORGANISM: murine
; US-08-628-145-7

Query Match      6.1%; Score 332; DB 2; Length 584;
Best Local Similarity 23.6%; Pred. No. 1.1e-19;
Matches 108; Conservative 108; Mismatches 190; Indels 52; Gaps 15;

QY 23 STLERQKPPPTPSGKLVSIQMLDDTQAEFEVQPARAFKVLDDAVCNHNLNVEGDYFG 82
Db 12 SSLKR--KQPKFTFVRIVM-----DAEMEFNCENKWKGDLDLVCRTLGLRDTWFFG 63

QY 83 LEFPDHKKITVLDLLKPIV-KQIRPKHVVKVFFVKKFPPD-HTQLQELTRYLFALQY 140
Db 64 LQY-TIKDTVAVLKMDDKVLHDVSKKEEPTVTHFLAKFYPENAEELVQVEITQHLFFLQV 122

QY 141 KDLAQGRLTCDNTSAALLSHIVQSEIGDFEAL-DREHLAKNKVIPOQ-----D 190
Db 123 KKEILDEKVCYCPPEASVLLASAVQAKYGDYDPSVHKRGFLAQEELPRVINLYQMTPE 182

QY 191 ALEDKIVEFHNNHIGQTPAESDFQLEIARRLEMVYGIHLHPAKDREGTKINLAVANTGIL 250
Db 183 MWEERITAWAHRGRARDEAEWYELKIAQDLEMVGVNYFTIRNKKGTETLLGLVDALGLH 242

QY 251 VFOGFTKIN---AFNWKVKRLSKFKRFLIKLRPDANSAYQDTLEFLMASRDFCKSWK 307
Db 243 IYDPENRLTPKISFPWNEIRNISYSDKFTIK----PLDKKIDVFKFDSKLRVKNKLIQ 298

QY 308 ICVEHHAFTRLPEEPKPKPVLFSGSSFRSGRTQKQVLDVYKGGHKKVQFERKHSK 367
Db 299 LCIGNHDLF--MRRRADSLVQOQKAQAEKARKOMERQRLARE---KOMREAEATR 353

QY 368 IHSIRLSAQPTL--LNSEVLEQSQOSTSLTFEGEAGSPGQSCRRGKPKVSAGEPGSH 425
Db 354 DELERRLLQMKKEATWANEALMRSEETADL-LAEKQAI-----TEEEAKLLAQKAAEA 405

QY 426 PSPAPRRSPAGNKQADGAASAPTEEEVVKDRTQOSK 463
Db 406 EQEMQRIK-----ATAIRTEEEKRLMEQKVLAE 434

```

Search completed: December 6, 2001, 08:55:08  
Job time: 290 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:46:28 ; Search time 44.17 Seconds  
(without alignments)  
1802.183 Million cell updates/sec

Title: US-09-555-342A-2  
Perfect score: 5463  
Sequence: 1 MGEIEQRPTGSLGAPENS.....SATSSASRPHVLHSHKESLVY 1045

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_68:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	5456	99.9	1045	JC5795	CDEP protein - hum
2	850.5	15.6	1111	T23047	hypothetical prote
3	709.5	13.0	850	MMHUE4	erythrocyte membra
4	698	12.8	926	A41105	protein-tyrosine-p
5	692	12.7	858	A46613	protein 4.1, P4.1
6	691.5	12.7	1698	T13800	coracine gene prote
7	686.5	12.6	230	T34541	hypothetical prote
8	681	12.5	801	A37353	membrane protein 4
9	644	11.8	1026	T19631	hypothetical prote
10	568	10.4	568	T25859	hypothetical prote
11	541	9.9	913	A41109	protein-tyrosine-p
12	475.5	8.7	554	J00188	membrane protein 4
13	465.5	8.5	414	T22498	hypothetical prote
14	457.5	8.4	4549	T20771	hypothetical prote
15	457.5	8.4	4667	T20774	hypothetical prote
16	441.5	8.1	961	A55380	factogenital dyspl
17	434	7.9	1175	S51005	protein-tyrosine-p
18	430	7.9	639	T20772	hypothetical prote
19	414.5	7.6	875	T19678	hypothetical prote
20	401.5	7.3	1176	T58345	protein tyrosine p
21	396	7.2	1174	T38140	protein-tyrosine-p
22	391.5	7.2	577	S39804	moesin - pig
23	390.5	7.1	577	A41289	moesin - human
24	386.5	7.1	583	A41129	radixin - mouse
25	385.5	7.1	583	S39805	radixin - pig
26	382.5	7.0	583	A46127	radixin - human
27	367.5	6.7	630	T47177	hypothetical prote
28	365	6.7	581	T45889	ezzrin - bovine
29	356.5	6.5	586	A34400	ezzrin [validated]

```

30 356 6.5 586 1 B41129
31 349 6.4 654 2 T30957
32 341.5 6.3 1187 1 JC4155
33 337 6.2 563 2 T29262
34 335 6.1 595 2 S33809
35 334 6.1 591 2 I54368
36 334 6.1 596 2 I68664
37 333.5 6.1 1189 1 JC2366
38 331 6.1 564 2 T29263
39 315 5.8 2490 1 A54971
40 300 5.5 559 2 A45620
41 299 5.5 559 2 S49143
42 296.5 5.4 2466 2 I67629
43 294.5 5.4 2294 2 I67630
44 293 5.4 2450 2 S71625
45 291 5.3 1737 2 A59235

ezrin - mouse
hypothetical prote
protein-tyrosine-p
hypothetical prote
neurofibromin 2 -
merlin protein - m
merlin - mouse
protein-tyrosine-p
hypothetical prote
protein-tyrosine-p
cytotoxin homolog
EG10 protein - tap
protein tyrosine p
protein tyrosine p
protein-tyrosine-p
unconventional myo

RESULT 1
JC5795
CDEP protein - human
C:Species: Homo sapiens (man)
C:Date: 24-Jan-1998 #sequence_revision 13-Mar-1998 #text_change 16-Jul-1999
C:Accession: JC5795
R:Koyano, Y.; Kawamoto, T.; Shen, W.; Yan, W.; Noshiro, M.; Fujii, K.; Kato, Y.
Biochem. Biophys. Res. Commun. 241, 369-375, 1997
A:Title: Molecular cloning and characterization of CDEP, a novel human protein contain
ng factors.
A:Reference number: JC5795; MUID:98086358
A:Accession: JC5795
A:Molecule type: mRNA
A:Residues: 1-1045 <KOY>
A:Cross-references: DBJ:AB008430
C:Comment: This protein is involved in the adhesion, proliferation, and differentiation
C:Superfamily: pleckstrin repeat homology; protein 4.1 membrane-binding domain homolo
F.1-374/Domain: ezrin-like #status predicted <E2R>
F.42-316/Domain: protein 4.1 membrane-binding domain homology <B41>
F.931-1027/Domain: pleckstrin repeat homology <PLK>

```

## ALIGNMENTS

```

Query Match 99.9%; Score 5456; DB 2; Length 1045;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1044; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MGEIEQRPTGSLGAPENSIGSTLERGQKPPPTPSGKLVSIKIQMLDDTQEAFAEVPQRA 60
DB 1 MGEIEQRPTGSLGAPENSIGSTLERGQKPPPTPSGKLVSIKIQMLDDTQEAFAEVPQRA 60

QY 61 PGKVLDDAVCNHNLNVEGDFYFGLFPDHHKKTITWLDLLKPIVKIRPKHVVVKKVVF 120
DB 61 PGKVLDDAVCNHNLNVEGDFYFGLFPDHHKKTITWLDLLKPIVKIRPKHVVVKKVVF 120

QY 121 PPDHTQLQELTRYLFALQVKQDLAQGRCLTNCNTSAALISHIVQSEIGDFDEALDREHL 180
DB 121 PPDHTQLQELTRYLFALQVKQDLAQGRCLTNCNTSAALISHIVQSEIGDFDEALDREHL 180

QY 181 ANKNTYPOQDALEDKIVFHHNHIGQTPAESDFOLLEIARRLEMVGIHLHPAKDREGTKI 240
DB 181 ANKNTYPOQDALEDKIVFHHNHIGQTPAESDFOLLEIARRLEMVGIHLHPAKDREGTKI 240

QY 241 NLAVANTGILVFGQFTKINAFNNAWKVRKLSFKRRKRLIKLRPDANSAYODTLEFLMASRD 300
DB 241 NLAVANTGILVFGQFTKINAFNNAWKVRKLSFKRRKRLIKLRPDANSAYODTLEFLMASRD 300

QY 301 FCKSFWKICVEHHAFFRLFEKPKPKPVLFSGSSFRFSGRTQKQVLDYVREGGHHKQY 360
DB 301 FCKSFWKICVEHHAFFRLFEKPKPKPVLFSGSSFRFSGRTQKQVLDYVREGGHHKQY 360

QY 361 FPKHSKIHISRLASLOPTELNSEVLEQSQOSTSLTFGGEAESPQSGSCRRKEPKVKSAG 420
DB 361 FPKHSKIHISRLASLOPTELNSEVLEQSQOSTSLTFGGEAESPQSGSCRRKEPKVKSAG 420

```

```

Db 361 PERKSHKIHSLRASQPTLNSEVLQSQOOSTSLTFEGEAESPGQOQRRGKPKVSAG 420
QY 421 EPSCSPSPARRSPAGNKQADGAASAPTEEEBWWKDRTOQSKPQPQSTGLTGSPL 480
Db 421 EPSCSPSPARRSPAGNKQADGAASAPTEEEBWWKDRTOQSKPQPQSTGLTGSPL 480
QY 481 SELSVNSQSGVAPANTYLSPLNLPDTPKQASPLISPLNDQACPTDDEGRKRKPTDK 540
Db 481 SELSVNSQSGVAPANTYLSPLNLPDTPKQASPLISPLNDQACPTDDEGRKRKPTDK 540
QY 541 AYPIAEKVESTERTYKLDLEVTISWFOSTVSKEDAMPEALKSLIFNPFPLHKHFTNFK 600
Db 541 AYPIAEKVESTERTYKLDLEVTISWFOSTVSKEDAMPEALKSLIFNPFPLHKHFTNFK 600
QY 601 EIFQRLALWGRSNAOIRYQRTGDVYMLKNIQGMKHLAHLAMKHSLEALENGIKSSRR 660
Db 601 EIFQRLALWGRSNAOIRYQRTGDVYMLKNIQGMKHLAHLAMKHSLEALENGIKSSRR 660
QY 661 LENFCRDFELQKVCYLPLNTFLLRPLRLMHYKQVLERLCKKHPPSHADFRDCRAALAEI 720
Db 661 LENFCRDFELQKVCYLPLNTFLLRPLRLMHYKQVLERLCKKHPPSHADFRDCRAALAEI 720
QY 721 TENVAQIHGTMKMFQKHELKLDLIGDNLVPGREFIRLGSLSKLSGKGLQORMEF 780
Db 721 TENVAQIHGTMKMFQKHELKLDLIGDNLVPGREFIRLGSLSKLSGKGLQORMEF 780
QY 781 LFNDVLLTYSRGLTASQKFGHLYGWTIEESEDWGVPHCLTLRGOROSIIVAASS 840
Db 781 LFNDVLLTYSRGLTASQKFGHLYGWTIEESEDWGVPHCLTLRGOROSIIVAASS 840
QY 841 RSEMERKVEDIQMAIDLAESKSPAPFLASSPPDNKSPDEATAADQESDDLSASRTSL 900
Db 841 RSEMERKVEDIQMAIDLAESKSPAPFLASSPPDNKSPDEATAADQESDDLSASRTSL 900
QY 901 EROAPHRGTMVHVCWHRTSVSMWDFSTAVENQLSGNLLRKFKNNGWQKLVVFTNFC 960
Db 901 EROAPHRGTMVHVCWHRTSVSMWDFSTAVENQLSGNLLRKFKNNGWQKLVVFTNFC 960
QY 961 LFYKSHQDNHPLASPLGLYSITIPSESENQKDVVKLFKSHVYYPRAESEYTFERW 1020
Db 961 LFYKSHQDNHPLASPLGLYSITIPSESENQKDVVKLFKSHVYYPRAESEYTFERW 1020
QY 1021 MEVIRSATSSASRPHVLSHKSLEY 1045
Db 1021 MEVIRSATSSASRPHVLSHKSLEY 1045

RESULT 2
T23047
hypothetical protein H05G16.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T23047
R:White, S.
submitted to the EMBL Data Library, June 1997
A:Reference number: Z19661
A:Accession: T23047
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1111 <WIL>
A:Cross-references: EMBL:Z97190; PIDN:CAB10024.1; GSPDB:GN000028; CESP:H05G16.1
A:Experimental source: clone H05G16
C:Genetics:
A:Gene: CESP:H05G16.1
A:Map position: X
A:Introns: 97/3; 139/2; 171/3; 213/2; 294/3; 348/2; 373/2

Query Match 15.6%; Score 850.5; DB 2; Length 1111;
Best Local Similarity 26.6%; Pred. No. 1.3e-41;
Matches 287; Conservative 157; Mismatches 363; Indels 271; Gaps 38;

QY 22 ISTLERGQKPPPTPS-----GKLVSIKIQMLDDTQEAPE----- 55

```

```

Db 1 MSNIPRGVAGAPPPMNSSKRGLVCIKVRMLDDTVAVFHLGVICFFCVVSPFHGYVEL 60
QY 56 VQORAPGKVLDAVCNHLNVECDYFGLFPHDKITVWLDLLKPIVKOIRR-PKHVVVK 114
Db 61 LQHKAIQTGLDDEVCHRLNLECDYFGLFIDINGNHCWLDREKTLTROITNGSTDAKY 120
QY 115 FVVKFPDPDHTOLOEELRYLFALOVKQDLAOGRLTCNDTSAALLSHIVQSEIGDF--D 172
Db 121 FVVKFPDPIDLEEBYTRYLFTMQIKRDLALGELHCSNDTASLLSAYLVQSECGDFS 180
QY 173 EALDREHLAKNKYIQQD-ALEDKIVFEFHHNIGOTPAESDFOLLEIARLEMYGIRLHP 231
Db 181 DYPDATYLSHTRFVNPOTLEFQKVMNDNRNFITGMTGESDLAMLEVARRCDFYGVKLA 240
QY 232 ADRECTKINLAVANTGILVFOGFTKINAFNNAKVRKLSFKRKRFLIKLRPDANSAYQDT 291
Db 241 AKDIDGNDAAALSVMLHGKIVFQLOLQDTTFSWARKLSFKRKKLLVLRHPSYQYLKET 300
QY 292 LEFLMASRDFCKSFWKICVEHHAFFRLPEPKPKPVLF-SRGSSFRSGRTQKQVLDY 350
Db 301 VFESFETRECKNFVKVEHHAFFRCVQAEPEKKEKTRFISGSSFRHGRTOQLIDY 360
QY 351 VKEGGHKKVOYFER----- 368
Db 361 VREHHKRRREPFTRLRSAASTRKGTYSSTYGLVTDRTKHRNGSVYEPNQTDPYNKHQNT 420
QY 369 HS-----IRLSAQPT-----LNSEVLEQ--SQOSTSLTFEGEAESPGQOQRRG-- 412
Db 421 HSMPIAHIISSQADHSGFGLDARVGSVTRDPSVTQHLRQLKRSERKISDVTV 480
QY 413 KSPKVSAGPGHSPAPRRSPAGNKQADGAASAP--EVEEVVKDRTOQSKPQPQPS 470
Db 481 ERQORSRCPVSHAINI--SSTSSKEQPDLSVLPVNSDDLOMVCKEIEIQNDPKSV 538
QY 471 TG-----SLTGSPLSELVNSOGGVAP--ANVTLSPLNLSPTDKQAS 510
Db 539 SGDNFOQRSSRDYDNVEDSYRLSDH--ERSTRSEVGVGSKFAAATIFNSTFVARPKGS 596
QY 511 PLISPLL-NDQACPRDDEDEGRK-----RPTDKAYFAKE----- 547
Db 597 NVVKRVVAHTKTPNSTDDDEGALKSASEYQTFRHIKEYPFARNANIVIEIDGNVDLS 556
QY 548 -----VSTERTYLKLEVTISWFOSTVSKEDAMPEALKSLIFN-PEPLHKHFTNLF 599
Db 657 ARRSPAATTTTTRTYANTTTTTSKVLTSAGVL--MKPKVISNDHEPSSGTHSS-- 711
QY 600 KEIQRALWEGRSNAOIRYQRTGDV-----MLKNIQGMKHLAAH----- 640
Db 712 -----RVSPESGTGALGPLGRVITKENMMITPEGFKEKKAKPNPPKPSFP 759
QY 641 -LWKHSEALEALENGIKSSRRLENFCRDFELQKVCYLPLNTFLLRPLRLMHYKQVLERL 699
Db 760 VOAVHSETVRE--IKTERASS-----HPLVHMQE--EIP 791
QY 700 CKHPPSHADFRDCRAALAEITEMVAQLHGTMIKMFNFKLHKLKDLIGDNLVVPGR 759
Db 792 FSRAPLKPEDKKERKSSLSRPLISVQSEDN---PDVKQKCHLFNSD-----IP--- 836
QY 760 FTRLGSLSKLKGLOORMFFLNDVLLYTSRG-----LTASQKFKVHQLPL---- 807
Db 837 ----YTLTMRNVENTQSLPYSSFKDVKSTAKQGYESNLSKRVSKSPEFKRRKSLDLPVR 892
QY 808 -----YGMTTESEDEGVPHCL----TLRGOROSIIVAASSRSEMEKKEWEDIQM 853
Db 893 RLPSGCFNSAQDHTTSTPTPDSVLEYLLRRSLLGDK-SVTKTKSRKDRPQRTQVPR- 950
QY 854 AIDLAESKSSPAPFLASSPP-----DNKSPDEATAADQESDDLSASRTSLERQAPH 906
Db 951 -FDLPPSPCSP--AGGSTPFISILNDOLFECVSESRLHEDMD----RLDKTTTPH 1000

RESULT 3

```



```

Db 660 SQEIKKHASISLKNFMESVPE-PRPSEWDKRLSTHSPFRFLNINQIPTGEGPLV 718
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 494 --ANVTLSPNLSPTKQASPLISLLNDQA-----CPRDD 527
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 719 KQTQVTISDNANAVKSEIPTKDVPIVHTETKTITYEAAQTDD 760
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

RESULT 4
A41105
protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 4 - human
N:Alternate names: PTPase MEG
C:Species: Homo sapiens (man)
C>Date: 20-Mar-1992 #sequence_revision 02-May-1994 #text_change 11-Jun-1999
C:Accession: A41105
R:Gu, M.; York, J.D.; Warshawsky, I.; Majerus, P.W.
Proc. Natl. Acad. Sci. U.S.A. 88, 5867-5871, 1991
A:Title: Identification, cloning, and expression of a cytosolic megakaryocyte protein-tyrosine-phosphatase
A:Reference number: A41105; MUID:91288564
A:Accession: A41105
A:Molecule type: mRNA
A:Residues: 1-926 <GUA>
A:Cross-references: GB:M68941; NID:gl90747; PIDN:AAA36530.1; PID:gl90748
A:Experimental source: megakaryocytes, cell line MEG-10
C:Genetics:
A:Gene: GDB:PTPN4
A:Cross-references: GDB:131387; OMIM:176878
A:Map position: 9q31-qg31
C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology; F
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatase
F:31-308/Domain: protein 4.1 membrane-binding domain homology <B41>
F:523-597/Domain: GLGF domain homology <GLG>
F:679-900/Domain: protein-tyrosine-phosphatase homology <PTP>
F:852/Active site: Cys (phosphocysteine intermediate) #status predicted
F:858/Binding site: substrate phosphate (Arg) #status predicted

Query Match 12.8%; Score 698; DB 1; Length 926;
Best Local Similarity 32.6%; Pred. No. 7.7e-33;
Matches 185; Conservative 89; Mismatches 198; Indels 96; Gaps 18;

QY 35 PSGKLVSIK-----IQMLDDTQEAPEVPQAPAGKVLDDAVCNHNLNVEG 78
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 8 PAGETYNVRASELARQDHTQEVVVCNILLNTVQAEKVNKHQDQGVLLVDFKHLDTQE 67
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 79 DYFLEPPDHKKITV-WLDDLLKPIVKQIRRPKHVVKVVFVPPDHTQAEELTRYLFA 137
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 68 DYFGLQADSTNPRWLDNPKPIRKQKRGSPYSLNFRVKFFVSDPNKLOEYTRYQYF 127
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 138 LQVKODLAQGRLTCDNTSAALLSHIVQSEIGDFDEALDRE-HLAKNKYIPQO-DALEDK 195
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 128 LQIKODILTLGRLPSPNTAALLASFAVQSELGVDYDQSENLSGVLSDFYFPNQPQDFEKE 187
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 196 IVEFHHNHIGTQPAESDFQLELTARLEMYGIRLHPAKDREGTKINLAVANTGILVFGQF 255
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 188 IAKLHQHQHIGLSPAEAFNTINTARTLELYGVFFHYARDQSNIEIMIGVMSGGILYKNR 247
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 256 TKINAFNNAVKRLSFKRKRFLIKLRDANSAYODTLEFLMASRDCKSPFKICVEHHAF 315
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 248 VRMTFFWLKIVKLSFKCKOFFQLRKELHESRETLGLGNVNYRACKNLWKAACVEHHTF 307
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 316 FRLEPEPKPKPKVYL---FSRGSFSRFGTKQKQVLDYVVEGGHKKVQFERKHSKTHSIR 372
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 308 FRL-DRELPPQNFFAHYFTLGSFKRYCGRTQVSQVQYQYGEKANKORVFARSPSK-PLAR 365
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 373 SLASQPTELNSVLEQSQQSTSTFTFGGAESPGQSGCRK--EPKVSAGEPGS----- 424
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 366 KLMDWEVYVSRNSISDDLETQSLP---SRSPPGTNNHRNSTFTQEGTRLRPSVGLHVD 421
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 425 ---HSPAP---RRSPAGNKQADGAA--SAPTEEE-----EEVVK 456
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 422 HMVHTSPSEVFNQSRPS-STQANSIVLESSPSQETPGDGKPPALPPKQSKKNSWNOIHY 480
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

```

```

QY 457 DRTOQ-----SKPQPPQSTGSLTGSPh-----LSELSVNSGGV 491
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 481 SHSQODLESHINETFDIPSSPEKPTPN---GGIPHDNLVLRMKPDENGRCFGFNKGGY 536
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 492 APANVTLSPNLSPTKQASPLISLLND 519
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 537 DQKMPVIVSRVAPGT--PADLCVPRLNE 562
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

RESULT 5
A46613
protein 4.1, p4.1 - mouse
C:Species: Mus musculus (house mouse)
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 25-Aug-1995
C:Accession: A46613
R:Huang, J.P.; Tang, C.J.; Kou, G.H.; Marchesi, V.T.; Benz Jr., E.J.; Tang, T.K.
J. Biol. Chem. 268, 3758-3766, 1993
A:Title: Genomic structure of the locus encoding protein 4.1. Structural basis for co
A:Reference number: A46613; MUID:93155238
A:Accession: A46613
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-858 <HUA>
A:Experimental source: embryo
A>Note: sequence extracted from NCBI backbone (NCBIN:124466, NCBIP:124467)
C:Superfamily: membrane protein 4.1; protein 4.1 membrane-binding domain homology
F:213-488/Domain: protein 4.1 membrane-binding domain homology <B41>

Query Match 12.7%; Score 692; DB 2; Length 858;
Best Local Similarity 36.0%; Pred. No. 1.5e-32;
Matches 179; Conservative 65; Mismatches 201; Indels 52; Gaps 12;

QY 5 EQRPPTGSRLL---GAPENSGISTLE----RGQKPP-----PTPSGKLVSIIKIQMLDDTQ 51
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 163 EHREDPDSETKEGEGTECGSTEVKEDPESRAEREPEASQKPVRRRNHMKVSLDDTV 222
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 52 EAEPQAPQAPGKVLDDAVCNHNLNVEGDYFGLFEPDHHKITVWLDLLKPIVKQIRR-PKH 110
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 223 YECVVEKHANLQDLKARVCEHLNLEEDYFGLALWDSATSKTWLDSAKELKKQVRGVPWN 282
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 111 VVVKVVFVPPDHTQAEELTRYLFAQVKQDLAQRGLTCNDTSAALLSHIVQSEIGD 170
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 283 FT--FNVKVEYPPDPAQLTEDITRYVLCQLQRIQDIVAGRLPCSPATLALLGSYTIQSELGD 340
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 171 FDEAL-DREHLAKNKYIPQO-DALEDKYIEFHNNHIGTQPAESDFQLELTARLEMYGIR 228
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 341 YDELHGMIDYVSDFKLAPNQTRELEEKVMELHKSYSKSMTPAQADLEFLENAKKLSMYGVD 400
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 229 LHPAKDREGTKINLAVANTGILVFGQFTKINAFNNAVKRLSPKRRKRFLIKLRPDANSAY 288
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 401 LHAKDLEGVDIILVCCSSGLLVYKDKLRINRPWPVKLVKISSKRSFFKIRPGQEHY 460
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 289 QDTLEFLMASRDCKSPFKICVEHHAFRLFEPEKPKPKPVLFVSRGSSFRFSRGTQKQVL 348
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 461 ESTIGFKLPSYAAKKLWKVCVEHHHTFFRL-TSTDITPKSKFLALGSKFRYSRGTQAQTR 519
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 349 DYYKEGGHKKVQFERKHSKTHSIRSLASQPTELNSVLEQSQOSTSLITFEGEGSPGGQS 408
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 520 QASALIDRPAPHFERTASK-----RASRLDGGAAAAEST-DRS 556
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 409 CRRGKPKVKSAGSPGSHSPAPRRSP-----AGNKQADGAAASAPTEEEVEVVKDR 459
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 557 PRPTSAPATAQSQVTEGPGRPKTKTPKEAVKVEEKGEEPAEPAPPEPT-EAWKVEKTH 615
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
QY 460 QQSKKPPQPPSTGSLFG 476
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:
Db 616 EVTVPTNSGDTQKLAG 632
      :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||: :||:

RESULT 6
TI3800
coracle gene protein - fruit fly (Drosophila melanogaster)

```

C:Species: Drosophila melanogaster  
C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 17-Nov-2000  
C:Accession: T13800  
R:Fehon, R.G.; Dawson, I.A.; Artavanis-Tsakonas, S.  
Development 120, 545-557, 1994  
A:Title: A Drosophila homologue of membrane-skeleton protein 4.1 is associated with septin  
A:Reference number: Z17763; MUID:94215495  
A:Accession: T13800  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-1698 <FEH>  
A:Cross-references: EMBL:L27467; NID:g440293; PID:g440294; PIDN:AA59187.1  
C:Genetics:  
A:Gene: coracle  
A:Cross-references: FlyBase:FBgn0010434  
A:Map position: 2  
C:Function:  
A:Description: may play a role in cell-cell interactions  
C:Keywords: alternative splicing

Query Match 12.7%; Score 691.5; DB 2; Length 1698;  
Best Local Similarity 28.9%; Pred. No. 4.5e-32;  
Matches 195; Conservative 107; Mismatches 237; Indels 139; Gaps 19;

QY	5	EQRTPGSRIGCAPENSIGISTLERGQKPPPTPS--GKLVSIKIQMLDDTQFAFVQPARG 62
DB	4	EIKPSAPAEPTPKS-----KPKSSSSSHGKPALARVTLLDGLSDVSDIRKAI 54
QY	63	KVLIDAVCNHLNIVEGDYFGLFEPDHHKIIVLWLLKPIVQIRPKHVVVVKVVFEP 122
DB	55	RDVINSICAGLNLEKDYFGLTYETDPTDPTWLDLEKPVSKFFRTDTPWLT-FAVKYI 113
QY	123	DHTLOQELRYLPALQVKODLAQRLTCNDTSAALISHIVQSEIGDFD--EALDREHL 180
DB	114	EPSQLKEDITRYHLCLVRNLDLEGLPCFVTHALLGSYLVQSEMGDYDAEEMPTRAYL 173
QY	181	AKNYIYQODA-LBDKIVFHHNIHGOTPAESDPQLLEIARRLEMYGIRLHPAKDRG 239
DB	174	KDFAIAPNOTAELDKYMDLHKTKGSPAEAEUHYLENAKKLAMYGVDLHPAKDSEGD 233
QY	240	INLAVANTGILVFOGTFKINAFNNAKVKLSFKRKRFLIKLRPDANSAYODTLEFLN 299
DB	234	IMLVGASGLLVYRDKLRINRFAPKILKISYKHHFYIKIRGPEFEQYESTIGFKLANH 293
QY	300	DFCKSWKICVEHHAFRLFEPEKPKPVLFSS-RGSSFRFSRGTQ----- 344
DB	294	RAAKLWKSCEVHHFTFFRLM-TPEVSKSMFVFGSTYRKRTQAESTNTPVDRTPPK 352
QY	345	-----KQVLDYVKEGHHKKVQFERKHSI-----HS-----IRSLASQP 378
DB	353	FNRTLSGARTSRSMALAAEKEV--ARKSSTLDHRGDNRADGDAHSRPIKNKKEKD 410
QY	379	TELNSEVLEOSQO-----STSLTEG--EGAESPGQSCRR 411
DB	411	ADKEAKLREKKQKEKEKERKEKRELEKKKAEKAAKALAGAAGAVNGDELND 470
QY	412	GKEPKVSAGEPGSHSPAPRRSPAGNKQADGAASAPTEEBEEV-----VKD 457
DB	471	SNKSDSGSRGVRGIFSSGRKSKSGSPKDGKDKSGDKDKDKEGRLGLVVTSGLDNQD 530
QY	458	RTQOSKQPPQPPSTGSLT-GSPHLSLSVNSOGVAPANTVLSL----- 500
DB	531	QNLDEAARNAAKNRGSTPGVTROYEAVDNDGNTSPTRKSYTPGGFRYDODPNRSKGA 590
QY	501	-----NLSPDTKK-----ASPLISPLNDQACPRTDDEDEGRKKRFFTDK 540
DB	591	DGQQLSPTSQCKKIGLVNAPGNNALNKAETAEKLAGQLSPQTQDK-LNRGQLSPKR 649
QY	541	AYTIAKE-VSTERTYLIK 557
DB	650	AKLLQDPLLSPPTTRAKLO 667

```

RESULT      7
T34541
hypothetical protein DKFZp434M221.1 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34541
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, October 1999
A:Reference number: Z21541
A:Accession: T34541
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-230 <BLU>
A:A/Cross-references: EMBL:AL12052
A:A/Experimental source: adult testis; clone DKFzP434M221
C:Genetics:
A>Note: DKFzP434M221.1

Query Match          12.6%; Score 686.5; DB 2; Length 230;
Best Local Similarity 63.8%; Pred. No. 4.7e-33;
Matches 132; Conservative 29; Mismatches 39; Indels 7; Gaps 3;

QY   831 ROSTIVAASRSSEMEKVVEDIQAIDLAEXSSPAPEF----LASPPDNKSPDEATAAD 886
DB    3 OKTIIVAASTRLKEKWMLDLSAIQAIAKSGGDTAPALPGTVCTRP--RSPNE-VSLE 59

QY   887 QESDDLSASRTSLERQAPHRGNTMVHVCVWHRNITSVMDFSIAVENQLSGNLRLRFKNS 946
DB    60 QESDDARGVRSLSLEGHGHRANTTMHVCVRYNITSVRADHSAAVENQLSGYLRLFKNS 119

QY   947 NGWKLVWFVTNFCFLPFYKSHQNPLIASIPLLGYSITIPSESENOKOVYVKLHKFSHV 1006
DB   120 HGWKLVWFVTNFCFLPFYKTHQDDIPLASILPLLGSVIPREADGIHKDYVFVKLFKSHV 179

QY   1007 YFRAESEYTFERMWEVIRSATSSASR 1033
DB   180 YFRAESKYTFERMWEVIQGASSAGR 206

RESULT      8
A37353
membrane protein 4.1 - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change 13-Aug-1999
C:Accession: A37353; A29901
R:Spencer, M.; Giebelhaus, D.H.; Kelly, G.M.; Bicknell, J.; Florio, S.K.; Milam, A.H.
Dev. Biol. 139, 279-291, 1990
A>Title: Membrane skeleton protein 4.1 in developing Xenopus: expression in postmitot
A:Reference number: A37353; MUID:90249600
A:Accession: A37353
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-801 <SPE>
A:A/Cross-references: GB:M20621; NID:g214090; PIDN:AAA49695.1; PID:g214091
R:Giebelhaus, D.H.; Fib, D.W.; Moon, R.T.
Cell 53, 601-615, 1988
A>Title: Antisense RNA inhibits expression of membrane skeleton protein 4.1 during em
A:Reference number: A29901; MUID:88223353
A:Accession: A29901
A:Molecule type: mRNA
A:Residues: 1-550 <GIE>
C:Superfamily: membrane protein 4.1; protein 4.1 membrane-binding domain homology
C:Keywords: cytoskeleton; membrane protein
F:195-470/Domain: protein 4.1 membrane-binding domain homology <B41>

Query Match          12.5%; Score 681; DB 2; Length 801;
Best Local Similarity 30.7%; Pred. No. 6.1e-32;
Matches 180; Conservative 80; Mismatches 212; Indels 114; Gaps 15;

QY   28 GQKPPTTPSGKL-----VSIKIQMLDQTQAEFVPQRAPCKVLLDVACVNHNLN175

```

Matches	188;	Conservative	82;	Mismatches	212;	Indels	84;	Gaps	20;
---------	------	--------------	-----	------------	------	--------	-----	------	-----

  

Qy	13	RIGAPENSIGITLBERGQPPPTPSGKLVISIKIOMLDLTQEAFEVQPAPRGKVLLDVCNH	72
Db	2	RLGSNSYDVQRTEAIGQTVPVKTPPNQIRCVTFLDSTSYHFEIEKNSGLGIVLLEKFVNY	61
Qy	73	LNLVEGDYFGLF-----PDHKKITWDLKLPIVKQ--IRRKHVVVKKVFVFFFPDH	124
Db	62	LEITEKDYFGLVFAVDNSSAQKK--WLDPSKNLRKQMICPPVHLF--PRVFYYVRDP	116
Qy	125	TQLQEELTRYLFAQVKODLAQGRLTCDNTSAALLISHIVQSEIGDPE---ALDREHLA	181
Db	117	NRLRDERTFOYQVORONLEEGRLPCNEGSLALLASVVOAEWGDFEETHGMSRCTLCLC	176
Qy	182	-KNKYIPOQDALEDKIVRFHNHHIGOTPAESDFOLLEIARLEMYGIRLHPAKDRGTKI	240
Db	177	YKIQFATLPDDFSDRAVELHQHIGQTFDVAEQNFDHARLEMYGMDVYDGV DANHLPI	236
Qy	241	NLAVANTGILVFGFTKINAFNAWKVRKLSPKRREFLIKLIPDANSAYQDTLFELMASRD	300
Db	237	EIGVGAVGIKFHEGIIKMNEYAWRIKLSPKKQFOV-LVANEDGVSETIMIFINSAK	295
Qy	301	FKCSFWKICVHHAFRLFEEPKPKPVLESRSSRRFSORTOKOVL---DYVKEGGHK	357
Db	296	ICKLLWKCCIEQHFTFRLLKTTPK-TPORKVFNFGSKRYSGRTYQTLREENHRKSAGHR	354
Qy	358	KVQPERKHSKTHSTRSLASOPTELNSEVLEQSQOSTSLTFCGEAESP-GGOSCRRGEPK	416
Db	355	--NHRSLSKSFLRSFTSG----NQSIDSRV-TNVTITDSLELSSGOLLAR---RL	404
Qy	417	VSAPEGSGHSPA-----PRRSPAGNKQAADGAASAAPTDEEVVKD	457
Db	405	LSAARHDTDSSDALGYASDGAVVCAPLTTPLSPRT--RDVATDSESSAPLSROORLSKE	462
Qy	458	R-----TQQS---KQOPPOPTGSLTGSPHLSLSV-----NSQGVA	492
Db	463	AIYGTQESCDEKSWTFTSMACSTSPGIHASTASVRPVSSGSTPNGASRKRSANSGYSGYG	522
Qy	493	PANVTLPNLSPDKQASPLISPLL N 518	
Db	523	YATQIQP-----YSTTNASYSPYL N 543	

RESULT 10

T25859

hypothetical protein T04C9.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T25859

R:Favellio, A.

submitted to the EMBL Data Library, December 1996

A:Description: The sequence of C. elegans cosmid T04C9.

A:Reference number: Z20101

A:Accession: T25859

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-568 <FAV>

A:Cross-references: EMBL:U80955; PIDN:AAB38104.1; GSPDB:GN00021; CBSP:T04C9.6

A:Experimental source: strain Bristol N2; clone T04C9

C:Genetics:

A:Gene: CESP:T04C9.6

A:Map position: 3

A:Introns: 52/3; 87/3; 129/2; 162/1; 197/2; 280/3; 393/3; 538/3

  

Query Match	10.4%	Score	568;	DB 2;	Length	568;			
Best Local Similarity	29.5%;	Pred. No.	1.4e-25;						
Matches	163;	Conservative	81;	Mismatches	205;	Indels	104;	Gaps	15;

  

Qy	38	KLVSIIKTMQLDDTQEAFEVQPAPRGKVLLDVCNHNLNVGEDYFGLFPPDHUKITVWLIDL	97
Db	33	KFVQCKVILLDGAHLNIIVPBNVAGSELVEYFYSLDIERDYFGLQYTFNPVQHMLDP	92

QY 98 LKPIVQIRPKHVYVKKFPDPD-HTQLOBELTRYFALQVQKODLAGRLTCNDTSA 156  
 Db 93 TRKVAQOVAIGPFTLRFRVYKFTSPSSNLKEELTRYOFFLQIKQIDSSGRLOCPHQLA 152  
 QY 157 ALLTSHVQSEIGDFEADLRE-HLAKNKYIPOOD-ALEDKIVEFHNNHIGQTPAESDFQ 214  
 Db 153 IELAAALQSELDGYNPETHALFISEFRPHODEKMEVEILE-----SGQTPAAQELN 207  
 QY 215 LLEIARLEMYGIRLHPAKDREGTKINLAVANTGIILVFOGFTKINAFNNAKVRKLSFKRK 274  
 Db 208 YLNKARWIEYMGVDMHIVEGCKDNTYRLGLTPQGLMLVFDGPOKIGLFLWEKQLKDFKNK 267  
 QY 275 RFLIKLRPDANSAYQ-----DTLEFLMASRDFCKSWKICVEHAFRFLFEKPKPKPV 329  
 Db 268 KITLVVEEDAQSNNGOIQIOLHTFVFLHTSEKAAKHFWMKAIEQHAFFRLKSRPIQANRKI 327  
 QY 330 LFSR-GSSFRFSQRTQKQVLDYVKEGCHKVQPERKHSKIHISIRSLASQPTELNSEVLEQ 388  
 Db 328 QFFRLGSTFKYRGTEYEYI-----HK----- 349  
 QY 389 SQOSTSLTFTGEAESPQSGCRGKPKVSAEPGSHPSAPRRSP----- 434  
 Db 350 -----EGARLSRRQSCSPERRPSQRYGPRQSHVTNAQIRDAKRAEMRQOIIIEQOR 399  
 QY 435 AGNKQADGAASAPTEEEV-----VKDRTQOSKP---QPPQSTGSLTGS---P 478  
 Db 400 ANDQQRTHAKSDPTSPHSIQNVPLPTSPDLKPPVNSNPFISDPLTPSSSSSSSHNLS 459  
 QY 479 HSELSVNSOGGV-----APANVTLSPNLSPDTKOASPLISPLLNDOA---CPRTDDE 528  
 Db 460 HVTKIIVGTIGNLDTSPSSNAFSSYIPRPVSTSTSESHP--PPLPAHQSNSKIPRMSSA 517  
 QY 529 DEGRKRKRPFTDKA 541  
 Db 518 SEQRKSSQPPQPA 530  
 RESULT 11  
 A41109  
 protein-tyrosine-phosphatase (EC 3.1.3.48), nonreceptor type 3 - human  
 N:Alternate names: PTPH1  
 C:Species: Homo sapiens (man)  
 C:Date: 27-Mar-1992 #sequence\_revision 02-May-1994 #text\_change 11-Jun-1999  
 C:Accession: A41109; 155698  
 R:Yang, Q.; Tonks, N.K.  
 A:Title: Isolation of a cDNA clone encoding a human protein-tyrosine phosphatase with ho  
 Proc. Natl. Acad. Sci. U.S.A. 88, 5949-5953, 1991  
 A:Reference number: A41109; MUID:91296738  
 A:Accession: A41109  
 A:Molecule type: mRNA  
 A:Residues: 1-913 <YAN>  
 A:Cross-references: GB:M64572; NID:g179912; PIDN:AAA35647.1; PID:g179913  
 R:Iikuta, S.; Itoh, F.; Hinoda, Y.; Toyota, M.; Makiguchi, Y.; Imai, K.; Yachi, A.  
 J. Gastroenterol. 29, 727-732, 1994  
 A:Title: Expression of cytoskeletal-associated protein tyrosine phosphatase PTPH1 mRNA i  
 A:Reference number: 155698; MUID:95179278  
 A:Accession: 155698  
 A:Status: translated from GB/EMBL/DBDJ  
 A:Molecule type: mRNA  
 A:Residues: 899-913 <RES>  
 A:Cross-references: GB:S76309; NID:g913165; PIDN:AAB33583.1; PID:g913166  
 C:Genetics:  
 A:Gene: GDB:PTPN3  
 A:Map position: 9q31-9q31  
 C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology; P  
 C:Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatas  
 F:31-508/Domain: protein 4.1 membrane-binding domain homology <B41>  
 F:516-590/Domain: GLGF domain homology <GLGF>  
 F:670-890/Domain: protein-tyrosine-phosphatase homology <PTP>  
 F:842/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:848/Binding site: substrate phosphate (Arg) #status predicted

RESULT 12  
 JU0188  
 membrane protein 4.1 homolog - mouse

Query Match 9.9%; Score 541; DB 1; Length 913;  
 Best Local Similarity 24.3%; Pred. No. le-23;  
 Matches 232; Conservative 150; Mismatches 346; Indels 228; Gaps 41;  
 QY 27 RGOKPPTSGKLVSIKIQMLDDTOFAFEVQPORAGKVLDDAVCNHLNLVEGDFYGLERP 86  
 Db 16 RTSSELPEKTRSERVICSIIHFDGVQVOTFKVTKQDQTQVLLDMVHNLGVTEKFEYGLQHD 75  
 QY 87 DHKKITV-WIDLKLPITVKQIRRPKHVVVVFVFFPDHTQLOEELTRYFALQVQKODLA 145  
 Db 76 DSDVSPLWLEASKPIRKQLKGGFPCTLHPVRFFIPDPNTLQOETRHLFLQLKMDIC 135  
 QY 146 QGRLTNDTSAAALLSHIVOSEITGDFDEALDRE-HLAKNKYIPOQDALED---KIVEFHH 201  
 Db 136 EGRLTCLPLASVAVLASYAVQSHFGYNSIIHPGYSLSHSFIPDON--EDFLKVESLHE 193  
 QY 202 NHTGQTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFOGFTKINAF 261  
 Db 194 QHSLGKQSEAESCIYINARTLDFYGVELHSGRDLHNLDMIGIASAGAVARYKICTSFY 253  
 QY 262 NNAKVRKLSFKRRKRLTIKLRPDANSAYQDTLEFLMASRDFCKSWKICVEHAFRFLFE 321  
 Db 254 PWNILKISFKRKKFTHORQKQAESREHIVAFNMLNYRSCKNLWKSVEHHTFFQA-KK 312  
 QY 322 PKPKPKPVL---FSRGSFRFSQRTQKQVLD-YVKE--GG-----HKVQPERKHSKI 368  
 Db 313 LLPQEKNVLSQYWTGMSR-----NTRKSVNNQYCKKYGIGVWVNPAMRRSLSVHELETK- 366  
 QY 369 HSTRSLASQPTELNSEVLEQSQSTSLTFTGEAESPQSGCRGKPKVSAEPGSHPS 428  
 Db 367 ----SLPSRSPPI-----TPNWR-----PR-----RHEIR 389  
 QY 429 APRRSAGNKAQADGAASAPT--EEEEVVKDRTQKQKPPQPPSTGSLTSGPHLSLSVN 486  
 Db 390 KPRHS-----SADNLANEMTYITETEDVFTYKGLAPQ-DSSEVVSQNSRSHQESLSEN 443  
 QY 487 SQGGVAPANVTLSPNLSPDTKOASPLISPLLNQOACPRDDEDEGRKRKRPPTKAYPIAK 546  
 Db 444 N-----PAQSYL-----TKSSSVSPSSNAPGSCPDGVQDQL-----LDDFHRVTK 486  
 QY 547 EVSTTERTYLDLEVTISFQSVSKEDAMPEALKSL-IPNFEPLKHKHTNFKLTEQR 605  
 Db 487 GGSTEDAS-----QYYCDKNDNGDSYLVLRITPDSD--GKGFNLKGGVDQK 532  
 QY 606 LALWEGRSNAQ-----IRDYQRTIGDVMKNIQGMKH-----LAAHLKHSLEALEA 650  
 Db 533 MPLVVSRIINPESPADTCIPKLNESDQIVLINGRDISEHTHDQVVMFKASRESHSREL-A 591  
 QY 651 LENGIKSSRRLENFCRDFELQKVCYILPLNTFLRLPLRLMHYKQVLERLCKHHPPSHADF 710  
 Db 592 LVIRRAVRASFADFKSEDELNLQ-----FPEAIFPMC-----PEGGD- 628  
 QY 711 RDCRAALAEITEMVAQIHGCTMIKWE-----NFOKLHELKLDLIGIDNLVVPVREF 760  
 Db 629 -----TEGSMALKGLSEGTVLIOFEQDYRKK-----PGLAI 662  
 QY 761 IRLGSLSKLSGKGLQORMFFLFNDVLLY-TSRGLTASNQPKVHG-----OLPLYGMTIEE 814  
 Db 663 ----TFAKLPQNLKDNK----YKDVLPYDTRVLLOGNEDYINASYVNMIEIPAANLVNKY 714  
 QY 815 SEDWGVPH-CL-----TLRGQROSIIVAASSRSEMEKWEKVEDIQMAIDLAESSSPAPEFL 869  
 Db 715 IATQGGPLPHTCAQFQWQVMDQLSLIV-----MLATLTERGRGRTKCHQYW 758  
 QY 870 ASSPPD-----NKSPDEATAAQSESDLLSASRTSLERQAPHRGNTMVHVCW 916  
 Db 759 -POPPDMNHGGPHIOCSSESDCTIAYVSRMLVNTQTGEEHTVTH-----LQYVAV 809



C:Species: Mus musculus (house mouse)  
C:Date: 03-Feb-1994 #sequence\_revision 20-Aug-1994 #text\_change 05-Nov-1999  
C:Accession: I55505; J00188  
R:Takeuchi, K.; Kawashima, A.; Nagafuchi, A.; Tsukita, S.  
J. Cell Sci. 107, 1921-1928, 1994  
A:Title: Structural diversity of band4.1 superfamily members.  
A:Reference number: I55505; MUID:95074267  
A:Accession: I55505  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-554 <RES>  
A:Cross-references: GB:D28818; NID:9466547; PIDN:BAA05978.1; PID:9466548  
C:Superfamily: protein 4.1 membrane-binding domain homology  
F:13-295/Domain: protein 4.1 membrane-binding domain homology <B41>

Query Match 8.7%; Score 475.5; DB 2; Length 554;  
Best Local Similarity 30.3%; Pred. No. 3.3e-20;  
Matches 135; Conservative 79; Mismatches 182; Indels 49; Gaps 14;  
QY 43 KIQMLDDTQAEFVQPR-----APGKVLDAVCNHLNVEGDYFGLFEPDPKHKITVWLDL 97  
DB 14 EVLLDESKTLTQOQGIKSKTSGVLDHVFRRHNLVEIDFGLRYCDRSHQTVLDP 73  
QY 98 LKPIVQK---IRPKHVYVVKFFPPDHTQLQOELTRYLFALQVQDLAQGLTCNDT 154  
DB 74 AKTLAEHKEKELINTGPPVTLFGIKFYAEDPCCKLEKITRYQFFLQVQDALQGLRCPVN 133  
QY 155 SAALLSHIVQSEIGDFDEALDRE-HLAKNKYIP-QQDALEDKIVFEHNNHICQTAE 212  
DB 134 IAAOMGAYATQAEGLDHPDKHTAGVSEYFVDPDQKEELEAEIRIKHTLMGQAPSEAE 193  
QY 213 FQLELTARLEMGIRLHPAKDREGTKINLAVANTGLVFOGFTKINAFNAKVRKLSFK 272  
DB 194 LNLVRLAKSLEMVGLDHPVVGNGENKSEYFGLTPSGVVVYTKNKKQVKYFWPRTKVHF 253  
QY 273 KRRLIK-LRPDANSAYQDTLEFLMASRDFCKSFVKICVEHHAFFRLEFEPKPKPVL 331  
DB 254 ETQELRVLGKDCN---ETSEFFFEARSKTACKHLWKCSVEHHTFFRMPDTESNLSRKLS 310  
QY 332 SRGS-SFRFSGRQKV-LDYVEGHHKKVQFERKISKTHSIRSLASQPTLSEVLEQS 399  
DB 311 KFGSISYKHYRTALQMSRDLSTQLPRPNQNVVRSKTYPKRVAQTQPTGSNN--INRI 368  
QY 390 QQSTSLTFEGG-----AESPGGSCRGKPKVSGAGEPGSHSPAPRRSPAGNKQAGAA 444  
DB 369 TANTENGEGTKITLAPSP-VKSFKKAK-----NENSPDPQRS-----KS 408  
QY 445 SAPTEEE-----EYVVKDRTOQSK 463  
DB 409 HAPWEENGPOSGLYNSSSDRTKSPK 433

RESULT 13  
T22498  
hypotheical protein F52D10.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 18-Feb-2000  
C:Accession: T22498  
R:Kershaw, J.  
submitted to the EMBL Data Library, November 1995  
A:Reference number: Z19571  
A:Accession: T22498  
A:Molecule type: DNA  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Residues: 1-414 <WIL>  
A:Cross-references: EMBL:Z66564; PIDN:CAA91472.1; GSPDB:GN000028; CESP:F52D10.6  
A:Experimental source: clone F52D10  
C:Genetics:  
A:Gene: CESP:F52D10.6  
A:Map position: X  
A:Introns: 46/2; 195/3; 264/1; 337/3; 389/2

Query Match 8.5%; Score 465.5; DB 2; Length 414;  
Best Local Similarity 28.0%; Pred. No. 8.2e-20;  
Matches 119; Conservative 87; Mismatches 176; Indels 43; Gaps 10;  
QY 623 IGDVMLKNTQGMKHLAAHLWKHSEA-LEALENGIKSRRLLENFCRDFELQKVCYLPLNTF 681  
DB 2 VAPVLLATVHNLTPLYQSLVQNFPYIVSALDQLYRSKIPFTIMSKFESKECYTQVNW 61  
QY 682 LLRPLRLMHYKOVLERLCKHHPPSHADPRDC-----RAALAEITEMVAQLHGTMI 732  
DB 62 LKILNRLINWQPVLRARVIE-----IQISECONDNETTAFGVAMDKILIEFAKTKATRQ 115  
QY 733 KMFNFOLKHEKMD--LIGIDNLVVPGRFIRLGSLSKLSGKLQOQRMFFLDFNLLYTS 790  
DB 116 SLEEYIHLVQVERDTGLVGI--LTHPNRKILRVGVFLRSARRAPCCCRIMVLCSDRILFGH 173  
QY 791 RGLTA-SNFKVHGQLPLVGMTIESEDEW---GVPHCLTLRGQROSIIVAASRSEMEK 846  
DB 174 RGVNLDGNFTVHAEPFKLGMIDEGDTYKVMGDNENNVITLHNADISIVFAAPDRA---S 230  
QY 847 WVEDIQMAIDLAEKSSPAPFLASSPPDNKSPDEATAADOESEDLSASRTSLERQAPH 906  
DB 231 WIEDITEAKNAARAKIDPLSLV-----MEKKEENDIDMTKVLPLESPV 275  
QY 907 RGNTHVHCWHRTSVSMVDFTSAVENQLSGNLLRKFKNKSGWKLWVVFTEFCLFFYKS 966  
DB 276 SKMSPLQICWYKCSFGRKDVNKMIITWCGYLKRLNRSNGHODLWVVMCCHTLYFYRN 335  
QY 967 HQNHPLASPLPLGYSLTTPSESENIOK-DYVEPKLHKSHVYFRAESETFFERMEVIR 1025  
DB 336 HNEREPALHLSMDYGVGLPTVADKIDNHENCKFLYGSHTYFRTDSYFFERWVDSIF 395  
QY 1026 SATSS 1030  
DB 396 QAAS 400  
RESULT 14  
T20771  
hypotheical protein ZK270.2a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T20771; T27814  
R:Kershaw, J.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19322  
A:Accession: T20771  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4549 <WIL>  
A:Cross-references: EMBL:Z81499; PIDN:CAB04090.1; GSPDB:GN000019; CESP:ZK270.2a  
A:Experimental source: clone F11C3  
R:Kershaw, J.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z20423  
A:Accession: T27814  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4549 <WIL>  
A:Cross-references: EMBL:Z82089; PIDN:CAB05003.1; GSPDB:GN000019; CESP:ZK270.2a  
A:Experimental source: clone ZK270  
C:Genetics:  
A:Gene: CESP:ZK270.2a  
A:Map position: 1  
A:Introns: 61/3; 112/1; 244/3; 328/2; 384/1; 406/3; 4302/3; 4343/3; 4377/2; 4469/3; 4

Query Match 8.4%; Score 457.5; DB 2; Length 4549;  
Best Local Similarity 28.1%; Pred. No. 8e-18;  
Matches 158; Conservative 83; Mismatches 239; Indels 83; Gaps 20;  
QY 28 GOKPPPTPS-----GKLVSIKIQMLDDTQAEFVQPRAGKVLDDAVCNHLNVEGDYFG 82



Db 27 GDQKHTTSQPRDSKQMLAKVLPDGVQKFEFVNKNSEGEALFRQVTRDLSIEERIFS 86  
Qy 83 LEFPDHHKIT-VWLDDLLKPIVKQIR-RPKHVVVVFVKKFFPPDHTQLQOELTRYLFALQV 140  
Db 87 LCFYDKDEGTRHWLYNDKNAQIKGLPWE--FSFEVKFYPTPTTIVDDHARYVFLQL 144  
Qy 141 KODLAQGRLLCNDTSAALLLSHIVOSIGDFD-EALD--REHLAKNKYIPOQDALED--- 194  
Db 145 RRDULTGRLLPATADTHSLGSLGTFVQIEFGDAPAEMTDAYEQFIVASKLVPSAQNPTYK 204  
Qy 195 KIVEFHNNHIGQTPAESDFOLLEIARLEMYGIRLHPA-KDREGTKINLAVANTGILVFQ 253  
Db 205 KIVDLHREMGRQTPSEAEQNFLDHCKHLYGIHLFKAISDKDKKPDVGVGGAAGINIYO 264  
Qy 254 GFTKINAFNAKVRKLSFKRKRRLIKLRPDANSAYQDTLEFLMASRDFCKSFVKICVEHH 313  
Db 265 DEQKTHFSQWNIKIGYRTYSIKUKAGTVEKNEKTYFLKPNHVAARTWKCAVEHH 324  
Qy 314 AFFRLEPEPKPKPVLFSGSS--FRPSGRTQKQVLDYVKEGGHKV--QFERKHSKHS 370  
Db 325 TFFRLI-QPEDKTHKSFNFSGRFRYQGRTOQT-----KIASQWFDKPSIVDR 373  
Qy 371 IRSLASOPELNSVLBSQSOQSTSLTGE-----GAESPQGOSCRRGKEPKVSAG 420  
Db 374 APSAMSQPI---ATAENQKQLTNLTDSLEQROFERYRRALTPKSYTSSRQDDAVSTA 430  
Qy 421 -----EPGSHPSAPARR---SPAGNKQADGAASAPTEEEEEVVKDRTOQSK 463  
Db 431 TFAKYSRPSRLVYSTSGHHQSPHELSYSPRGDSSNSYSAAYMSESSL--RTPSSA 488  
Qy 464 PQPQPST-----GSLTSPHLSVNSOGGVAPANVTLSPNL-----SPD 505  
Db 489 YYPSEVTSPTSPYVFEGQHGEYHVQMRTASTSATGSPNSGARRNLFGRSDARSSQA 548  
Qy 506 TKQASPLIS---PLLNDAQCPRT 525  
Db 549 SRDSRLVSPHEPVDNPNEPIPT 571

RESULT 15  
T20774  
hypothetical protein ZK270.2d - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T20774; T27818  
R:Kershaw, J.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19322  
A:Accession: T20774  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-4667 <WIL>  
A:Cross-references: EMBL:Z81499; PIDN:CAB54224.1; GSPDB:GN00019; CESP:ZK270.2d  
A:Experimental source: clone F11C3  
R:Kershaw, J.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z20423  
A:Accession: T27818  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-4667 <WIL>  
A:Cross-references: EMBL:Z82089; PIDN:CAB54513.1; GSPDB:GN00019; CESP:ZK270.2d  
A:Experimental source: clone ZK270  
C:Genetics:  
A:Gene: CESP:ZK270.2d  
A:Map position: 1  
A:Introns: 61/3; 112/1; 244/3; 328/2; 384/1; 406/3; 4302/3; 4343/3; 4384/1; 4422/3; 4463/1

Query Match 8.4%; Score 457.5; DB 2; Length 4667;  
Best Local Similarity 28.1%; Pred. NO. 8.3e-18;  
Matches 158; Conservative 83; Mismatches 239; Indels 83; Gaps 20;

Search completed: December 6, 2001, 08:52:34  
Job time: 366 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:46:48 ; Search time 28.69 Seconds  
(without alignments)  
1335.474 Million cell updates/sec

Title: US-09-555-342A-2

Perfect score: 5463

Sequence: 1 MGEIEQRPTGSRIGAPENS.....SATSSASRPVLSHKESLVY 1045

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	708.5	13.0	864	1 4L_HUMAN	P11171 homo sapien
2	698	12.8	926	1 PTH4_HUMAN	P29074 homo sapien
3	694	12.7	881	1 E4L1_HUMAN	Q9H490 homo sapien
4	692	12.7	858	1 4L_MOUSE	P48193 mus musculus
5	681	12.5	801	1 4L_XENLA	P11434 xenopus lae
6	644	11.8	1026	1 PTP1_CAEEL	P28191 caenorhabdi
7	626	11.5	732	1 YF48_HUMAN	Q9Hcm4 homo sapien
8	541	9.9	913	1 PTN3_HUMAN	P26045 homo sapien
9	532	9.7	619	1 NBL4_BRARE	O57457 brachydanio
10	518.5	9.5	598	1 NBL4_HUMAN	Q9Hcs5 homo sapien
11	475.5	8.7	554	1 NBL4_MOUSE	P52963 mus musculus
12	450.5	8.2	960	1 FGD1_MOUSE	P52734 mus musculus
13	441.5	8.1	961	1 FGD1_HUMAN	P98174 homo sapien
14	434	7.9	1175	1 PTNL_RAT	Q62728 rattus norv
15	401.5	7.3	1176	1 PTNL_MOUSE	Q62136 mus musculus
16	396	7.2	1174	1 PTNL_HUMAN	Q16825 homo sapien
17	391.5	7.2	576	1 MOES_MOUSE	P26041 mus musculus
18	391.5	7.2	576	1 MOES_PIG	P26042 sus scrofa
19	390.5	7.1	576	1 MOES_HUMAN	P26038 homo sapien
20	386.5	7.1	583	1 RAD1_MOUSE	P26043 mus musculus
21	385.5	7.1	583	1 RAD1_PIG	P26044 sus scrofa
22	382.5	7.0	583	1 RAD1_HUMAN	P35241 homo sapien
23	365	6.7	580	1 EZR1_BOVIN	P31976 bos taurus
24	356.5	6.5	585	1 EZR1_HUMAN	P15311 homo sapien
25	356	6.5	585	1 EZR1_MOUSE	P26040 mus musculus
26	341.5	6.3	1187	1 PTNE_HUMAN	Q15678 homo sapien
27	335	6.1	595	1 MERL_HUMAN	P35240 homo sapien
28	334	6.1	596	1 MERL_MOUSE	P46662 mus musculus
29	333.5	6.1	1189	1 PTNE_MOUSE	Q62130 mus musculus
30	327	6.0	572	1 MOES_LYTV	P52962 lytechinus
31	327	6.0	1051	1 YC94_HUMAN	Q9P242 homo sapien
32	308	5.6	2485	1 PTND_HUMAN	Q12923 homo sapien
33	302	5.5	578	1 MOES_HUMAN	P46150 drosophila

#### RESULT 1

ID	4L_HUMAN	STANDARD;	PRT;	864 AA.
AC	P11171; P11176; Q9Y578; Q9Y579; Q14245;			
DT	01-JUL-1989 (Rel. 11, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	PROTEIN 4.1 (BAND 4.1) (P4.1) (EPB4.1).			
GN	EPB41 OR E41P.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (NON-ERYTHROID FORM).			
RX	MEDLINE=88234496; PubMed=3375238;			
RA	Tang T.K., Leto T.L., Correas I., Alonso M.A., Marchesi V.T.,			
RA	Benz E.J. Jr.;			
RT	"Selective expression of an erythroid-specific isoform of protein			
RT	4.1.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 85:3713-3717(1988).			
RN	[2]			
RP	SEQUENCE FROM N.A. (NON-ERYTHROID FORM).			
RX	MEDLINE=89132003; PubMed=3223413;			
RA	Tang T.K., Leto T.L., Marchesi V.T., Benz E.J. Jr.;			
RT	"Expression of specific isoforms of protein 4.1 in erythroid and non-			
RT	erythroid tissues.";			
RL	Adv. Exp. Med. Biol. 241:81-95(1988).			
RN	[3]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 378-393 (ERYTHROID ISOFORM).			
RC	TISSUE=Reticulocytes;			
RX	MEDLINE=87092279; PubMed=3467321;			
RA	Conboy J., Kan Y.W., Shohet S.B., Mohandas N.;			
RT	"Molecular cloning of protein 4.1, a major structural element of the			
RT	human erythrocyte membrane skeleton.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 83:9512-9516(1986).			
RN	[4]			
RP	SEQUENCE FROM N.A. (ISOFORM 3).			
RX	MEDLINE=91217063; PubMed=2022644;			
RA	Conboy J.G., Chan J.Y.C., Chasis J.A., Kan Y.W., Mohandas N.;			
RT	"Tissue- and development-specific alternative RNA splicing regulates			
RT	expression of multiple isoforms of erythroid membrane protein 4.1.";			
RL	J. Biol. Chem. 266:8273-8280(1991).			
RN	[5]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RA	Huang S.C., Wang C., Lichtenauer U., Vortmeyer A., Zhuang Z.;			
RT	"Sequence of protein 4.1 from a human neuroblastoma cell line: LAN5.";			
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
RN	[6]			
RP	SEQUENCE OF 157-227 FROM N.A., AND VARIANT ILE-214.			
RA	Lichtenauer U., Huang S.C., Vortmeyer A., Zhuang Z.;			
RT	"Valine to isoleucine polymorphism in exon 4 of human protein 4.1.";			
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.			
RN	[7]			
RP	SEQUENCE OF 648-714.			
RX	MEDLINE=87008553; PubMed=3531202;			

#### ALIGNMENTS

RA Correas I., Speicher D.W., Marchesi V.T.;  
 RT "Structure of the spectrin-actin binding site of erythrocyte protein  
 RL 4.1";  
 RN J. Biol. Chem. 261:13362-13366(1986).  
 [8]  
 RP PHOSPHORYLATION AT TYR-660.  
 RX MEDLINE=91271361; PubMed=1647028;  
 RA Subrahmanyam G., Bertics P.J., Anderson R.A.;  
 RT "Phosphorylation of protein 4.1 on tyrosine-418 modulates its  
 function in vitro";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:5222-5226(1991).  
 [9]  
 RP CARBOHYDRATES.  
 RX MEDLINE=90036892; PubMed=2808371;  
 RA Inaba M., Maede Y.;  
 RT "O-N-acetyl-D-glucosamine moiety on discrete peptide of multiple  
 protein 4.1 isoforms regulated by alternative pathways.";  
 RL J. Biol. Chem. 264:18149-18155(1989).  
 CC -I- FUNCTION: PROTEIN 4.1 IS A MAJOR STRUCTURAL ELEMENT OF THE  
 ERYTHROCYTE MEMBRANE SKELETON. IT PLAYS A KEY ROLE IN REGULATING  
 MEMBRANE PHYSICAL PROPERTIES OF MECHANICAL STABILITY AND  
 DEFORMABILITY BY STABILIZING SPECTRIN-ACTIN INTERACTION. BINDS  
 WITH A HIGH AFFINITY TO GLYCOPHORIN AND WITH LOWER AFFINITY TO  
 BAND III PROTEIN.  
 CC -I- ALTERNATIVE PRODUCTS: 6 ISOFORMS; 1 (SHOWN HERE), 2, 3, ERYTHROID,  
 NON-ERYTHROID A AND NON-ERYTHROID B; ARE PRODUCED BY ALTERNATIVE  
 SPLICING.  
 CC -I- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES  
 AND EACH PHOSPHORYLATION EVENT SELECTIVELY MODULATES THE PROTEIN'S  
 FUNCTIONS.  
 CC -I- PTM: PHOSPHORYLATION ON TYR-660 REDUCES THE ABILITY OF 4.1 TO  
 PROMOTE THE ASSEMBLY OF THE SPECTRIN/ACTIN/4.1 TERNARY COMPLEX.  
 CC -I- PTM: O-GLYCOSYLATED; CONTAINS N-ACETYLGLUCOSAMINE SIDE CHAINS  
 IN THE CARBOXYL-TERMINAL DOMAIN.  
 CC -I- DISEASE: DEFICIENCY OF E4P1 IS A CAUSE OF HEREDITARY  
 ELLIPTOCYTOSIS (HE) AND OF HEREDITARY PYROPOIKILOCYTOSIS (HPP).  
 CC -I- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; J03796; AAA35793.1; -  
 DR EMBL; J03796; AAA35794.1; -  
 DR EMBL; M14993; AAA35795.1; -  
 DR EMBL; M61733; AAA35797.1; -  
 DR EMBL; AF156225; AAD42222.1; -  
 DR EMBL; AF156226; AAD42223.1; -  
 DR PIR; A60244; MMHUL4.  
 DR PIR; A26656; MMHUE4.  
 DR GlycosuiteDB; P11171; -  
 DR MIM; 130500; -  
 DR MIM; 266140; -  
 DR InterPro; IPR000299; Band\_4.1.  
 DR Pfam; PF00373; Band\_41; 1.  
 DR PRINTS; PR00935; BAND41.  
 DR SMART; SM00295; B41; 1.  
 DR PROSITE; PS00660; BAND\_41\_1; 1.  
 DR PROSITE; PS00661; BAND\_41\_2; 1.  
 DR PROSITE; PS00557; BAND\_41\_3; 1.  
 KW Structural protein; Alternative splicing; Cytoskeleton;  
 KW Phosphorylation; Pyropoikilocytosis; Glycoprotein; Elliptocytosis;  
 KW Polymorphism.  
 FT DOMAIN 264 420 BAND 4.1-LIKE.  
 FT DOMAIN 488 614 HYDROPHILIC.  
 FT DOMAIN 615 714 SPECTRIN-ACTIN-BINDING.  
 FT DOMAIN 715 864 CARBOXYL-TERMINAL.  
 FT MOD\_RES 660 660 PHOSPHORYLATION (BY EGFR).  
 FT VARSPLIC 1 209 MISSING (IN NON-ERYTHROID ISOFORM B).

FT VARSPLIC 228 262 ERYTHROID ISOFORM AND ISOFORM 3).  
 FT MISSING (IN NON-ERYTHROID ISOFORM A AND  
 FT NON-ERYTHROID ISOFORM B).  
 FT MISSING (IN ERYTHROID ISOFORM, NON-  
 FT ERYTHROID ISOFORM A, NON-ERYTHROID  
 FT ISOFORM B AND ISOFORM 2).  
 FT VARSPLIC 635 648  
 FT VARSPLIC 649 669 MISSING (IN ISOFORM 3).  
 FT MISSING (IN NON-ERYTHROID ISOFORM A AND  
 FT NON-ERYTHROID ISOFORM B).  
 FT VARSPLIC 772 805 MISSING (IN ERYTHROID ISOFORM).  
 FT VARIANT 214 214 V -> I.  
 FT /FTID=VAR\_009122.  
 FT CONFLICT 51 51 Q -> H (IN REF. 5).  
 FT CONFLICT 76 76 S -> N (IN REF. 5).  
 FT CONFLICT 168 168 F -> S (IN REF. 6).  
 FT CONFLICT 259 259 A -> T (IN REF. 5).  
 FT CONFLICT 665 665 N -> S (IN REF. 5).  
 FT CONFLICT 679 679 K -> E (IN REF. 5).  
 FT CONFLICT 802 802 K -> Q (IN REF. 4 AND 5).  
 SQ SEQUENCE 864 AA; 97016 MW; B4731249D7FBE31 CRC64;  
 Query Match 13.0%; Score 708.5; DB 1; Length 864;  
 Best Local Similarity 32.2%; Pred. No. 4.2e-33;  
 Matches 192; Conservative 87; Mismatches 224; Indels 93; Gaps 18;  
 QY 16 APENSGISTLERGOKPPPTSPGKLVSIKIQMLDDTQEAPEVQORAPGKVLDDAVCHNLN 75  
 DB 188 SPQSKAETELKASQK--PIKRNHMKCKVSLDDTVYECVVEKHAQODLLKRVCEHLN 245  
 QY 76 VEGDYFGLFEPDHKKITVWMLDLKPIVKQIRR--PKHVVKVYVFPFPPDHTQLOEELRY 134  
 DB 246 LEEDYFGLAIDWNATSKTWLDSAKEIKKQVGVVWNT--FNVKYPDPDPAQLTEDITRY 303  
 QY 135 LFLAQVKQDLAQLRITCNDSALLSHIVQSEIGDFDEALDR--EHLAKNKYIPQ--DAL 192  
 DB 304 YLCQVLRQDIVAGLPCSFATLALGYSYIQSELGVDYDPDELHGVYDVKFLAPNQTKEL 363  
 QY 193 EDKIVFEHNNHIGOTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGLVLF 252  
 DB 364 EKVYMLHKSYSRMTQAADLEFLNAKLSMGVDLHRAKDLGVDIILGVCSSGLLYV 423  
 QY 253 QGFTKINAFNNAVKRLSPKRRFLIKLRPDANSAYODTLEFLMASRDFCKSPWKICVEH 312  
 DB 424 KKLRLNRPWPVKLVKISYKRSFFKIRPGEQYESTIGFKLPSYRAAKKLWKVCVEH 483  
 QY 313 HAFRLFEFPKPKPVLFSGSSFRSGRTQKQVLDYVKEGGHKVQPERKHSKHSIR 372  
 DB 484 HTFFRL--TSTDITPKSKFLALGSKFRYSRGTQATQATROASALIDRPAHPHFERTASK--RASR 541  
 QY 373 SLASQPTELNSEVLEQSQOSTSLTFEGEASPGG---QSCRRGKPKV-----SA 419  
 DB 542 SLGAAAVDSADRS PRPTSAPAITQGVAE--GGVLDASAKKTVPKAKETVKAEVKKE 599  
 QY 420 GEPGSHSPAPRRS-----PAGNKQADGAASAPTEE----- 450  
 DB 600 DEPPEQAEPEPTAMKVKETHIEVTPTNSGDTQKLAETEDLIRMKKKRRLDGENI 659  
 QY 451 -----EEV-----VKDRTOOSKPOPPPOPSGSLTGSPPH--LSELS 484  
 DB 660 YTRHSLMLEDLKSOEEIKKHHASISELKNFMESVPE--PRPSEWDKRLSTHSPFRTLN 718  
 QY 485 VNSQ--GGVAP-----ANVTLSNLSPTKQASPLISPLINDQA-----CPRTDD 527  
 DB 719 INQIPTGEGPPLVKVTQVTTISDNANAVKSEIPTKDVPIVHTETKTITYEAAQTDD 774  
 RESULT 2  
 ID PTN4 HUMAN  
 AC P29074; STANDARD; PRT; 926 AA.  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)







```
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Lightning J., Thomas K.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 844-950 FROM N.A.
RX MEDLINE=91139172; PubMed=1704870;
RA Matthews R.J., Flores E., Thomas M.L.;
RT "Protein tyrosine phosphatase domains from the protochordate Styela
RT plicata.";
RL Immunogenetics 33:33-41(1991).
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z36237; CAA85272.1; -.
DR EMBL; Z48241; CAA85272.1; JOINED.
DR EMBL; Z48241; CAA88287.1; -.
DR EMBL; Z36237; CAA88287.1; JOINED.
DR EMBL; M38013; AAA28127.1; -.
DR HSSP; P28827; 1RPM.
DR WormPep; C48D5.2; CE17578.
DR InterPro; IPR000299; Band_4.1.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR000387; Tyr_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00373; Band_41; 1.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTPHPPTASE.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00194; Ptpc; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00557; PDZ; 1.
DR PROSITE; PS00106; PDZ; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase; Cytoskeleton.
FT DOMAIN 87 245 BAND 4.1-LIKE.
FT DOMAIN 777 1026 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT_SITE 952 952 BY SIMILARITY.
FT SEQUENCE 1026 AA; 115093 MW; 2F43F7A614EDBC59 CRC64;
SQ
Query Match 11.8%; Score 644; DB 1; Length 1026;
Best Local Similarity 33.2%; Pred. No. 2.6e-29;
Matches 188; Conservative 82; Mismatches 212; Indels 84; Gaps 20;
QY 13 RLCAPENSGLSLERQKPPPTSGKLVIKIQMLDDTQEAFFVQAPQAPCKVLLDAVCNH 72
Db 2 RLGSNSYDVQRTAIGQTPVKTPPPNQIRCTVTFELDSYVHFEIKNLSGLVLEKVFNY 61
QY 73 LNLVEGDYFGLPEP-----PDHKRTVMLLDKPTVKQ-IRRPKRVVVKVFPFPPDH 124
Db 62 LEIENDYFGLVFIADVNSSAQKK---WLDPSKRLRKQMCIPPYHLF--FRVKFVRDP 116
QY 125 TQLOEBELTYFLAQVKQDLAQGRLCNTDTSALLSHIVQSIGDFDE----ALDREHLA 181
```

```
Db 117 NLRDEFTRFQFYQQVRQNLNLEGRPCNEGSLALLASYVYVQAEVGFEEKTHGMSRTCLC 176
QY 182 -KNKIYPOQDALEDKIVEPHNHIGOTPAESDFQLEIARRLEMYGIRLHPAKDREGTKI 240
Db 177 YKIQPATLPDDFSRVAELHQLHIGQTPDVAEQNFIDHARRLEMYGMDVYDGVDAHLPI 236
QY 241 NLAVANTGILVFQGGFTKINAFNNAKVRKLSFKRRLIKLRPDANSAYQDTLEFLMASRD 300
Db 237 EICGVAGKIVKPEHIGIKMNEYAMVRKLSFKKKQFQV-LVANEDGVSETIMIFNIMSAK 295
QY 301 FCKSFVKIVVEHAFPLFEPEPKPKPVLFSGSSFRFSGRGKQVLF---DIVKEGGHK 357
Db 296 ICKLLAKKCCIEQHTFLKTPPK-TPOKVFNFSGFRYSGRTEYQTLSENEHRKSAGHR 354
QY 358 KVQFERKHSKIHRSLSASQPTLNSEVLQSOOSTSLTEGGAESP-GQSCRRRGKEPK 416
Db 355 --NFHRSLSKSSFLRSTFGS---NTQSIDSSRY-TNTTTTSPDPELPSSQQLLAR---RL 404
QY 417 VSAGEPGSHPSPA-----PRRSAGNKQADGAASAPTEEEVEVKD 457
Db 405 LSAARHDTSSDALGYASDGAAGVVCAPLITPLSPRRT--RDYATDESSAPSLRQRLSKE 462
QY 458 R-----TQOS---KPQPPQSTGSLTSPHLSV-----NSQGGVA 492
Db 463 AIYGTQESCKDEKSWTPSMACTSTSPGIIHASTASVRPVSSGSTPNGASRKANSYSGYG 522
QY 493 PANVTLSPLNLSPTDKOASPLISPLN 518
Db 523 YATQTOOP-----TSTTNASYSYPLN 543
RESULT 7
YF48.HUMAN
ID YF48.HUMAN STANDARD; PRT; 732 AA.
AC Q9HCM4; Q9H975;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA1548.
GN KIAA1548.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Aotsuka S., Yoshikawa Y.,
RA Matsunawa H., Ishii S., Kawai Y., Saito K., Yamamoto J., Wakamatsu A.,
RA Nakamura Y., Nagahara K., Masuho Y., Sasaki N.;
RT "NEO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 188-732 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=20450683; PubMed=10997877;
RA Nagase T., Kikuno R., Nakayama M., Hirokawa M., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes.
RT XVIII. The complete sequences of 100 new cDNA clones from brain which
RT code for large proteins in vitro.";
RL DNA Res. 7:273-281(2000).
CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AK023019; BAB14360.1; -.
```





DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 1.  
 KW Structural protein; Cytoskeleton; Hydrolase.  
 FT DOMAIN 84 241 BAND 4.1-LIKE.  
 FT DOMAIN 510 582 PDZ.  
 FT DOMAIN 670 913 PROTEIN-TYROSINE PHOSPHATASE.  
 FT ACT\_SITE 842 842 BY SIMILARITY.  
 SQ SEQUENCE 913 AA; 104029 MW; 29A539ACDE2F1515 CRC64;

Query Match 9.9%; Score 541; DB 1; Length 913;  
 Best Local Similarity 24.3%; Pred. No. 1.7e-23;  
 Matches 233; Conservative 150; Mismatches 346; Indels 228; Gaps 41;

QY 27 RGQKPPPTSGKLSIRIQMLDQAEFVQAPAGKVLDDAVCNHNLNVEGDYFGLEFP 86  
 DB 16 RTSELPREKTRSEVICSIFHFDGVQFVKYQDGTQGVLLDMVHNLGVTEKEVGLQHD 75  
 QY 87 DHKITY-WDLLKPIYKQIRRPKHVVVVFVFPDPHTQLOEELRYLFAVQKODLA 145  
 DB 76 DSDVSPRWLEASKPIRKQLKGGPCTLHFRVRFPIPDNTLQEQTRHLYFLQKMDIC 135  
 QY 146 OGRITCNDTSAALLSHIVQSEIGDFDEALDRE-HLAKNKYIPQDALED---KIVEFHH 201  
 DB 136 EGRITCPLNSAVVLASAVOSHFGDYNSSIHHPGYSLSHFIPDN--EDFLTKVESLHE 193  
 QY 202 NHIGQTPAESDFQLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFQGTKINAF 261  
 DB 194 QHSLGKQSEAEVCINTARTLDYFVGLHSGRDLNLDLMIGTASAGVAYRYKICTSFY 253  
 QY 262 NNAVKRLSKRRFLIKLRPDANSQVDTLEFLMASRDFCKFWKTCVBEHHAFFRLEE 321  
 DB 254 PWNILKISFKRRKFFTHQKQAESREHVFANMLNYSKCNLWKSCEVHHHTFFQA-KK 312  
 QY 322 PKPKPKVL---PSRGSSFRSGRTQKQVLD-VYKE---GG-----HKVQPERKHSKI 368  
 DB 313 LLPQKKNVLSQYWTMGR-----NTRKSVNNQYCKKVIIGVGNWNPAMRRSLSEVHELETK- 366  
 QY 369 HSIRLASQPTELNSELVLEQSQSTSLTFGEAESPGQSCRRGKPKVAGEPGSPSP 428  
 DB 367 ----SLPSRPPI-----TPNWR-----PRL-----RHEIR 389  
 QY 429 APRSPAGNKQADGAASAPT--BEERVVKDRTQOSKQPQPQSTGLTSGPHLSLSVN 486  
 DB 390 KPRHS-----SADLANEMTYITETEDVFTYKGLSAPQ--DSDSEVSQNRSPHQSLEN 443  
 QY 487 SQGVAPANTVLSNLSPTDKQASPLISPLNDQACPTDDEGRKRKREPTDKAYFIAK 546  
 DB 444 N-----PAQSYL-----TQKSSSVSPSSNAPGSCSPGVQQL-----LDDEHRTVK 486  
 QY 547 EVSTERTYKDLVITSWFQSTVSKEDAMPEAKSL-IPNFEPLHKFTNFKLEIQOR 605  
 DB 487 GGSTEDAS-----QYQCDKNDGSDVLLVLRITPDDE--GKFGNLKGGVDQK 532  
 QY 606 LALWEGSRNAQ-----IRDYQIRGDMVNLKNIQMKH-----LAAHLWKHSEALEA 650  
 DB 533 MPLVSRINPESPADTCIPKLNEDGQVLVINGRDISHTHDQVVMFIFKASRESHREL-A 591  
 QY 651 LENGIKSRRLNFCRDFELQKCYLPLNTFLRLPLRLHMYKQVLERLCKHHPHSHADP 710  
 DB 592 LVIRRAVRFAFPKSEDELNL-----PPEATFPMC-----PEGGD- 628  
 QY 711 RDCRAALAEITEMVAOLHGTMIKME-----NFQKLHELKLDLIDNLVWPGREF 760  
 DB 629 -----TLEGSMQALKGLESCTVLIQFQELRYKK-----PGLAI 662  
 QY 761 IRLGSLKSLSGLOQRMFFLNDVLLY-TSRGLTASNQFVHG-----QLPLYGMTIEE 814  
 DB 663 -----TFALKQNLKDN-----YKDLVPDYTRVLQGNEDYINASYNNMVEIPANLVNKY 714  
 QY 815 SEDEWGVPH-CL-----TLRGQROSIIVASSRSEMEKWEVDIOMADLAEKSSSPAPEFL 869  
 DB 715 IATQGPLPHTCAQFQWVWQKLSLIV-----MLTTLTERGRTCKHQYV 758

QY 870 ASSPPD-----NKSPDEATAADQESDDLSASRTSLERQAPHRGNTMVHVCM 916  
 DB 759 -PDPPDVNMHGHFIQCSQSDCTIAYVSRMLVNTQTGEHTVTH-----LQTVAM 809

RESULT 9  
 ID NBL4\_BRARE STANDARD; PRT; 619 AA.  
 AC 057457;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DE 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE BAND 4.1-LIKE PROTEIN 4 (NBL4 PROTEIN).  
 GN NBL4.  
 OS Brachydanio rerio (Zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;  
 OC Cypriniformes; Cyprinidae; Rasbora; Rasbora; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE-98212541; PubMed-9551184;  
 RA Kelly G.M., Reversade B.;  
 RT "Characterization of a cDNA encoding a novel band 4.1-like protein in zebrafish";  
 RL Biochem. Cell Biol. 75:623-632(1997).  
 CC -!- FUNCTION: NOT KNOWN, BINDS CALMODULIN.  
 CC -!- TISSUE SPECIFICITY: IN ADULTS, IT IS FOUND IN THE OVARY, EYE, HEART, AND BRAIN, BUT NOT IN GUT OR SKELETAL MUSCLE.  
 CC -!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF025306; AAB97965.1; .  
 DR ZFIN; ZDB-GENE-990415-20; nbl4.  
 DR InterPro; IPR000299; Band\_4.1.  
 DR Pfam; PF00373; Band\_41; 1.  
 DR PRINTS; PR00935; BAND41.  
 DR SMART; SM00295; B41; 1.  
 DR PROSITE; PS00660; BAND\_41\_1; 1.  
 DR PROSITE; PS00661; BAND\_41\_2; 1.  
 DR PROSITE; PS00057; BAND\_41\_3; 1.  
 KW Structural protein; Cytoskeleton.  
 FT DOMAIN 70 230 BAND 4.1-LIKE.  
 SQ SEQUENCE 619 AA; 70708 MW; 0DFD7324DE513663 CRC64;

Query Match 9.7%; Score 532; DB 1; Length 619;  
 Best Local Similarity 29.4%; Pred. No. 3.1e-23;  
 Matches 146; Conservative 83; Mismatches 178; Indels 90; Gaps 14;

QY 43 KIQMLDD-----TQAEFVQAPAGKVLDDAVCNHNLNVEGDYFGLEFPDHKKITVWLDL 97  
 DB 14 EVLLDESKLLITQOQGIKKSTRGSVLDYVFSHVNLAETEYFGVRYCDRSHQFWLDP 73  
 QY 98 LKPIVKQ---IRRPKHVVVVFVFPDPHTQLOEELRYLFAVQKODLAQGRITCNDT 154  
 DB 74 SKTLAEHKDLIATGPPTLYFGVKFYAEDPGKLEETRYOFFLQVQKQVLDQRLPCAFN 133  
 QY 155 SAALLSHIVQSEIGDFDEALDRE-HLAKNKYIP-QDALEDKIVEFHHNHQTPAESD 212  
 DB 134 ISQAALAIQSELGSDYDPYKHTAGVSEYRFPVDFQKEDLEDSEIQHITLLQGVPAEE 193  
 QY 213 FQLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFQGTKINAFNNAVKRLSKFK 272  
 DB 194 NNLATAKTLEMVGVLDLHPVGEKQAEYFLGLTPGVVVVYKNTQVCKYFWPRITKYFK 253



CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----  
CC EMBL; D28818; BAA05978.1; -  
CC MGD; MGI:103007; Ebb4.114.  
CC InterPro: IPR000299; Band 4.1.  
CC Pfam: PF00373; Band 41; 1.  
CC PRINTS; PR00935; BAND41.  
CC SMART; SM00295; B41; 1.  
CC PROSITE; PS00660; BAND 41.1; 1.  
CC PROSITE; PS00661; BAND 41.2; 1.  
CC PROSITE; PS50057; BAND 41.3; 1.  
KW Structural protein; Cytoskeleton.  
FT DOMAIN 8 231 BAND 4.1-LIKE.  
SQ SEQUENCE 554 AA; 64082 MW; 7E63BAD088DBEBE CRC64;

Query Match 8.7%; Score 475.5; DB 1; Length 554;  
Best Local Similarity 30.3%; Pred. No. 4.5e-20;  
Matches 135; Conservative 79; Mismatches 182; Indels 49; Gaps 14;  
QY 43 KIQMLDDTQFAFVPPQ-----APGKVLDAVCNHLNVLVEGDFGLFFPDHKKITVWLDL 97  
Db 14 EVLLDESCLTLTQQGKIKKSGSVLDVHFRHNLVEIDYFGLRYCDRSHQTYWLDP 73  
QY 98 LKPIVQ---IRPKHVVKVFFPPDHTQLOELTRYLFALQVQDLAQRGLRCNDT 154  
Db 74 AKTAEHKLINTGPPYTLFYGKFAEDPCKLEITRYQFFLVQVQDALQRLPCPVN 133  
QY 155 SAALLSHVQSEIGDFDEALDRE-HLAKNKYIP-QQDALEDKIVEFHNNHIGOTPAESD 212  
Db 134 IAQMGAIALQELGDDHDPKHTAGVSEYRFPDQKEELEAERIHKTLMQGFSEAE 193  
QY 213 FOLLEIARLEMYGIRLHPAKDREGTKINIAVANTGILVFGQFTKINAFNWAQRKLSEK 272  
Db 194 LNLVLRATKSLMGVLDLHPYVGENKSEYFGLTSPGVVYKKNQVQKYEWPRTIKVHF 253  
QY 273 KRFLIK-LRPDANSAYQDTLEFLMSARDCKSFWKICVCEHHAFFRLFEPEKPKPVLF 331  
Db 254 ETQFELVLVQKDCN---ETSFEEARSKTACKHLMKCSVEHHFFFRMPDPSNLSRKL 310  
QY 332 SRGS-SPRFSGRTQKOV-LDYVVEGGHKVQFERKSHKSHIRSLASOPTELNSEVLEQS 389  
Db 311 KFGSISYKHYRTALQMSRDLSTQLPRPNQNVRSRKYIPKVAOTQPTGSNN--INRI 368  
QY 390 QQSTSLTFGR-----AESPGQSCRRGKPKVKSAGEPGSHPSAPRRSPAGNKQADGAA 444  
Db 369 TANTENGENGTKIITAPSP-VKSFKKAK-----NENSPDPQRS-----KS 408  
QY 445 SAPTEEE-----EVVVKDRTQSK 463  
Db 409 HAPWEENGQSGLYNSSSDRTKSPK 433

RESULT 12  
ID FGD1\_MOUSE STANDARD; PRT; 960 AA.  
AC P52734;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)  
DE (FACIOGENITAL DYSPLASIA PROTEIN HOMOLOG).  
GN FGD1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE=96081343; PubMed=8535076;  
RA Pasteris N.G., de Gouyon B., Cadle A.B., Campbell K., Herman G.E.,  
RA Gorski J.L.;  
RT Cloning and regional localization of the mouse facio-genital  
RT dysplasia (fgdi) gene.;  
RL Mamm. Genome 6:658-661(1995).  
CC BY EXCHANGING BOUND GDP FOR FREE GTP.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 FIVE-TYPE ZINC FINGER.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U22325; AAA96001.1; -  
CC MGD; MGI:104566; Fgdi.  
CC InterPro: IPR001849; PH.  
CC InterPro: IPR000219; RhoGEF.  
CC InterPro: IPR000822; Znf-C2H2.  
CC InterPro: IPR000306; Znf\_FYVE.  
CC Pfam; PF01363; FYVE; 1.  
CC Pfam; PF00169; PH; 2.  
CC Pfam; PF00621; RhoGEF; 1.  
CC SMART; SM00064; FYVE; 1.  
CC SMART; SM00325; RhoGEF; 1.  
CC PROSITE; PS50178; ZF\_FYVE; 1.  
CC PROSITE; PS50003; PH\_DOMAIN; 2.  
KW Guanine-nucleotide releasing factor; Zinc-finger.  
FT DOMAIN 7 330 PRO-RICH.  
FT DOMAIN 171 179 SH3-BINDING (POTENTIAL).  
FT DOMAIN 179 187 SH3-BINDING (POTENTIAL).  
FT DOMAIN 589 688 PH 1.  
FT ZN\_FING 729 789 FYVE-TYPE.  
FT DOMAIN 820 920 PH 2.  
SQ SEQUENCE 960 AA; 106477 MW; 41C1B84DE490FC51 CRC64;

Query Match 8.2%; Score 450.5; DB 1; Length 960;  
Best Local Similarity 22.9%; Pred. No. 2.6e-18;  
Matches 189; Conservative 131; Mismatches 338; Indels 169; Gaps 29;  
QY 320 EEPKPKPVLFSGSGSRFSGRTQKQVLDYVVEGGHKVQFERKSHKSHIRSL----- 374  
Db 142 ETPGPRSP-----LKRACPKPQVPKPSYLOMPVLPPPEP 179  
QY 375 ----ASQPTLNSEV-----LEQSQSTSLTFGEGASPGQSCRRGKPKVSAG----- 420  
Db 180 IPPPSRPLPADPRVAKGLVPRAEASTS-----SAVSSSLIEKPER--EPVIVASDRPAG 233  
QY 421 -----EPGSHPSAPRRSPAGNKQADGAAS-----APTEEEVVKDRTQ-----SKP 464  
Db 234 PCVPPEPAMLPPQPPQ--PTGSQPEGEASRCLFLAPGPRGEGKVPNRRSDSIDISSP 291  
QY 465 QPPQPTSGTLTSGPHLSVNSQGVAPANVTLSPLNSPDTKQASPLSLPLNDQACPR 524  
Db 292 SNSEETCFVSDGPPHSLCPG-----PALASMPVALADPHRPGSQEVDSDLEEE----- 342  
QY 525 TDDEDEGRKRFP-----DKAYFIAKEVSTERTYLKDLVITSWFQSTV 570  
Db 343 EEEEEKEEREIPVPPMERQESVELTVQKVFHIANELLQTEKAYVSRLLHLLDQVFCARL 402  
QY 571 SKE-----DAMPEALKSLIFNPPELKHFTNF--LKEIEQRLALWEGRSNAQIRDYQRIGD 625  
Db 403 LEEARNRSFPADVHGFISNICSIIYCFHQQLLPLEKRMEDW-----RYPRIGD 454

QY 626 VMLKNTQGMKHLAAHLKHSLEALENGIKSSRRLENFCRDFELQKVC-YLPNTLTLR 684  
Db 455 ILQKLAPFLKMYGEYKFNDRVAVELVNTWTSTQPKVIIHEVQKEACRNLTFLQHHMLE 514  
QY 685 PLRLHMYKQVRLKCKHHPSPSHADFCRAALAEITVMVAQLHGTWIKMKNFQKHLK 744  
Db 515 PVQIRPYELLKLDYLLKLPKPHGSPDSKDAKSLLELTATAEHSHNAIRKMERHKLKVVY 574  
QY 745 KDLIGDNLVVPREFIRIGLSKLSKSGK--LQORMFFLNDVLLYTSRGLTASNO-FKV 801  
Db 575 ELLGGEEDIVPTKELIKHGLKLSAKNGTQDRYLLILFNDRLLYCVPRLLGQKFTV 634  
QY 802 HGQPLPLYGMTIESEDEWGVPHCLTLRGQRSIIIVASRSSEMEKVVEDIQMAIDLAES 861  
Db 635 RARIDVGMELKESN-LNMPRTFLVSGKRSLEQARTTEEEKKDWQAINSTL-LKHEQ 692  
QY 862 SSPAPEFLAS-----SPDNKSPD---EATAADOESDDL---SASRTSLERQAPH- 906  
Db 693 TLETFKLLNSTNRDDETPPNSPNVDLGRKAPPIREKEVTMCMRQEPFNSITKRRHHC 752  
QY 907 -----RGNTMVHVCW-----HRNTSVSMV 925  
Db 753 KACGHVVCGKSEFARLIYDNNKSRNCTDCVVALHGAPGSSPACSOHTPQRRRSILEK 812  
QY 926 DFSIAVENQSLGNLKRK-FKNSNGWOKLWVFTN---FCLFFYKSHQDNHPLASLPLGY 981  
Db 813 QASVAEANSVICSFLHYMEKSGKGGWKAWEVVPNEPLVLIYCAPQDVKAQKRSPLIGF 872  
QY 982 SLTIPSESENIQDYVKFLHFKSHV-YFRAESEYTFERMEVIRSA 1027  
Db 873 EVGPPEAGERDRRHVFKI-TQSHLSWYSPETEELQRRMVAVLGRA 918

RESULT 13  
FGD1\_HUMAN  
ID FGD1\_HUMAN STANDARD; PRT; 961 AA.  
AC P98174;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PUTATIVE RHO/RAC GUANINE NUCLEOTIDE EXCHANGE FACTOR (RHO/RAC GEF)  
DE (FACIOGENITAL DYSPLASIA PROTEIN).  
GN FGD1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Craniofacial;  
RX MEDLINE=95042764; PubMed=7954831;  
RA Pastoris N.G., Cadle A., Logie L.J., Porteous M.E.M., Schwartz C.E.,  
RA Stevenson R.E., Glover T.W., Wilroy R.S., Gorski J.L.,  
RT "Isolation and characterization of the facio-genital dysplasia  
(Aarskog-Scott syndrome) gene: a putative Rho/Rac guanine nucleotide  
exchange factor.";  
RT Cell 79:669-678(1994).  
RL  
CC -!- FUNCTION: ACTIVATES THE RAS-LIKE FAMILY OF RHO- AND RAC PROTEINS  
CC BY EXCHANGING BOUND GDP FOR FREE GTP.  
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).  
CC -!- TISSUE SPECIFICITY: FETAL HEART, BRAIN, LUNG, KIDNEY, PLACENTA,  
CC AND LESS IN LIVER; ADULT HEART, BRAIN, LUNG, SKELETAL MUSCLE, AND  
CC LESS IN PANCREAS AND LIVER.  
CC -!- DISEASE: FACIOGENITAL DYSPLASIA (AARSKOG-SCOTT SYNDROME) IS A RARE  
CC MULTISYSTEMIC DISORDER CHARACTERIZED BY DISPROPORTIONATELY SHORT  
CC STATURE, AND BY FACIAL, SKELETAL, AND UROGENITAL ANOMALIES.  
CC -!- SIMILARITY: CONTAINS 1 DEL-HOMOLOGY DOMAIN (DH).  
CC -!- SIMILARITY: CONTAINS 2 PH DOMAINS.  
CC -!- SIMILARITY: CONTAINS 1 FIVE-TYPE ZINC FINGER.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: U11690; AAA57004.1; -  
CC MIM: 305400; -  
DR InterPro: IPR001849; PH.  
DR InterPro: IPR000219; RhoGEF.  
DR InterPro: IPR000822; Znf-C2H2.  
DR InterPro: IPR000306; Znf\_FYVE.  
DR Pfam: PF01363; FYVE; 1.  
DR Pfam: PF00169; PH; 2.  
DR Pfam: PF00621; RhoGEF; 1.  
DR SMART: SM00064; FYVE; 1.  
DR SMART: SM00233; PH; 2.  
DR SMART: SM00325; RhoGEF; 1.  
DR PROSITE: PS0178; ZF\_FYVE; 1.  
DR PROSITE: PS50003; PH\_DOMAIN; 2.  
KW Guanine-nucleotide releasing factor; zinc-finger.  
FT DOMAIN 7 330 PRO-RICH.  
FT DOMAIN 171 179 SH3-BINDING (POTENTIAL).  
FT DOMAIN 179 187 SH3-BINDING (POTENTIAL).  
FT DOMAIN 590 689 PH 1.  
FT ZN\_FING 730 790 FIVE-TYPE.  
FT DOMAIN 821 921 PH 2.  
SQ SEQUENCE 961 AA; 106668 MW; CF722598853A685A CRC64;

Query Match 8.1%; Score 441.5; DB 1; Length 961;  
Best Local Similarity 23.7%; Pred. No. 8.6e-18;  
Matches 173; Conservative 112; Mismatches 299; Indels 147; Gaps 26;

QY 413 KEPVSAGE---PGSHPS-----PAPRRSPAGNKQADGAAS-----APTEEEEEVVK 456  
Db 220 REPVIASDRPVGPGSPGPPVNLQPTSQPPVQPLPEGEASRCLEFLAAGPRDGEKVP 279  
QY 457 DRTQQ---SKQPPQ-----PSTGSLTGSPLHSELVSNQGVAPANVTLSPNL 502  
Db 280 NRDSGIDSISSPSNSETCFVSDGPPSHSLCPGPAL-----ASVPVALADPH 328  
QY 503 SPDTKQASPLISLLNDQACPRTDEDEGRK--RPT-----DKAYFIK 546  
Db 329 RPSQEV-----DSLEEDDEDEEEKRETPVPLMERQESVELTVQKVFHIAN 379  
QY 547 EVSTERTYTLKLEVITSWFQSVSK-----DAMPEALKSLIFPNFPLHKFTNF-LKE 601  
Db 380 ELLQTEKAYVSRLLHLLDQVFCARLLEARNRSSFADVVHGFISNCSIVCFHQQLLPE 439  
QY 602 IEQRLALWEGRSNAQIRYQIRGDMVKNTQGMKHLAAHLKHSLEALENGIKSSRRL 661  
Db 440 LKRMEEWD-----RYPRIGDILQKLAPFLKMYGEYKFNDRVAVELVNTWTSTQF 491  
QY 662 ENCRFQELQKVC-YLPNTLTLRPLRLHMYKQVRLKCKHHPSPSHADFCRAALAEI 720  
Db 492 KVIIEHVQKEACGNLTQHHMLPQVQIRPYELLKLDYLLKLPKPHGSPDSKDAKSLLEI 551  
QY 721 TEMVAQLHGTWIKMKNFQKHLKHLKDLIGDNLVVPREFIRIGLSKLSKSGK--LQORM 778  
Db 552 ATAEHSNAAIRKMERHKLKLVYELLGGEEDIVPTKELIKHGLKLSAKNGTQDRY 611  
QY 779 FFLFNDVLLYTSRGLTASNO-FKVHQLPLYGMTIESEDEWGVPHCLTLRGQRSIIIV 837  
Db 612 LILFNDRLLYCVPRLLGQKFSVRARIDVDMELKESN-LNLPRTFLVSGKRSLELQ 670  
QY 838 ASSRSEMEKVVEDIQMAIDLAESSSPAPEFLAS-----SPDNKSPD---EATAAQ 887  
Db 671 ARTEEEKKDWQAINSTL-LKHEQTLTFTKLLNSTNRDDETPPNSPNVDLGRKAPPIR 729  
QY 888 ESEDDL---SASRTSLERQAPH-----RGNTMVHVCW--- 916  
Db 730 EKEVTMCMRQEPFNSITKRRHCKACGHVVCGKSEFARLIYDNNKSRNCTDCVVAL 789

```
QY 917 -----HRNTSVSMVDFSIIVNOLSGNLLRKE-KNSNGMQLWVFTN-- 958
      | : | | | : | : | | | | | | | | | | | | | | | |
Db 790 HGVPGSSPACSQHTPQRRSILEKQASVAENSVICSLFHYMEKGGKGMKAWFVVPENE 849
      | : | | | | | | | | | | | | | | | | | | | | |
QY 959 -FCLFFYKQSHDNHPLASLPLGYSITPSENIQKDVFKLHFKSHV-YFRASEY 1016
      | : | | | | | | | | | | | | | | | | | | | | |
Db 850 PLVLYIGAPQDVKAQSLPLIGFEVGPPEAGERDRRHVKI-TQSHLSWYFSPETEEL 908
      | : | | | | | | | | | | | | | | | | | | | | |
QY 1017 FERWMEVIRSA 1027
      | | | | | | | | | | | | | | | | | | | | | |
Db 909 QREWMAVLGRA 919

RESULT 14
PTNL_RAT PTNL_RAT STANDARD; PRT; 1175 AA.
AC Q62728; Q62732;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 21 (EC 3.1.3.48)
DE (PROTEIN-TYROSINE PHOSPHATASE 2E).
GN PTPN21 OR PTP2E.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=95104449; PubMed=7805871;
RA L'Abbe D., Banville D., Tong Y., Stocco R., Masson S., Ma S.,
RA Fantus G., Shen S.H.;
RT Identification of a novel protein tyrosine phosphatase with sequence
RT homology to the cytoskeletal proteins of the band 4.1 family.;
RL FEBS Lett. 356:351-356(1994).
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + PHOSPHATE.
CC -1- TISSUE SPECIFICITY: PARTICULARLY ABUNDANTLY IN ADRENAL GLANDS.
CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U17971; AAA62153.1; -
CC EMBL; U18293; AAA62154.1; -
CC HSSP; P18031; 1PTT.
CC InterPro; IPR000299; Band_4.1.
CC InterPro; IPR000387; Tyr_phosphatase.
CC InterPro; IPR000242; Tyr_prot_phptase.
CC Pfam; PF00373; Band_41; 1.
CC Pfam; PF00102; Y_phosphatase; 1.
CC PRINTS; PR00700; PRTYPHPHTASE.
CC PRINTS; PR00935; BAND41.
CC SMART; SM00295; B41; 1.
CC SMART; SM00194; PTPC; 1.
CC PROSITE; PS00660; BAND_41_1; 1.
CC PROSITE; PS00661; BAND_41_2; 1.
CC PROSITE; PS00057; BAND_41_3; 1.
CC PROSITE; PS00383; TYR_PHOSPHATASE.1; 1.
CC PROSITE; PS00055; TYR_PHOSPHATASE.2; 1.
CC PROSITE; PS00056; TYR_PHOSPHATASE.2; 1.
KW Structural protein; Cytoskeleton; Hydrolase; Alternative splicing.
FT DOMAIN 77 241 BAND 4.1-LIKE.
FT DOMAIN 922 1175 PROTEIN-TYROSINE PHOSPHATASE.
```

```
FT ACT_SITE 1109 1109 BY SIMILARITY.
FT VARSPIC 1 839 MISSING (IN ISOFORM 2E).
SQ SEQUENCE 1175 AA; 133411 MW; 82A684F1C0F5ECF7 CRC64;

Query Match 7.9%; Score 434; DB 1; Length 1175;
Best Local Similarity 21.8%; Pred. No. 3.1e-17;
Matches 258; Conservative 171; Mismatches 458; Indels 298; Gaps 49;

QY 25 LERQKPPPTPSKIVSIKIQLDDTQEAFAFVQAPQAPVLLDAVCNHNLNVEGDYFGL 84
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 9 LKTRTRYTVSSKSLVA-RIQLLNNEFVETLSTVESTQESLEAVQAQRLRELTITYFSL 67
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 85 EPDHKKITVWLDLLKPIVKOIRPKHV--VKFVKFPPPDHQTQOEELTRYLLEALOVK 141
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 68 YYNQNRWVDLEKPLKQL--DKHALEPTVYGVVFPVSVSQQLQELTRYLYLQLK 125
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 142 QDLAQGLTCNDTSAALLISHIVQSEIGDFDEALDREHLAKNKYIP-----QDALED-- 194
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 126 KDVLEGNLPCTLEQAQLAGLAVQADFGDFQVESQDFLQKFALLPVGMQLDEKLEEA 185
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 195 -KIVEPHHNIGOTPAESDFQLEIARRLEMYGIRLHPAKDREGTKINLAVANTGILV-- 251
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 186 QKVALLHQYRGLTAPEAEMLYMQEYERMDGYEESYPAKDSQSDISIGACLDGIFVKH 245
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 252 FQGETKINAFNNAKVKLSFKRKRFLIKLPDANSAYQDTLEFLMASRDFCKSFWKICVE 311
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 246 KNGRPV-VFRHDIAMSHNKSFFALEL---ANK--EETIQOTEDMETAKYVWLICVA 299
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 312 HHAFRL-----FEPPKPK-----PKPVLSFGSSFRFSGRTQ 344
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 300 RHKFLRLNQLQTQAATLNSVRRGSSRSLPKPKQPYAMPPPQLHYNGHYTEFPASSQ 359
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 345 KQVLDVVKSG--CHKVKQFER-----KHSKIHISRS-----LASQP 378
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 360 DNVEFVNKNGFYCHSQTSLDRTQIDLSGRIRNGSVYSAHSTNSLNTFPQYLPQSPMSNP 419
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 379 TELNSVLE---QSOOSTSLTFGEAGSPGQS---CRRG-----412
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 420 SIPIGSDVMPDPIPSHRHSALIPPSYRPTPDYESVMKLRNGMVHADRHSHSLRNIGS 479
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 413 -----KEPKVSAGEPGSH-----SPAPRRSPAGNKQADGAASA 446
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 480 SVAYSRPDALVYSQPIREHPHLASQSAHYFPNLNVSFHSQAPYVPVRRPVVGAVSV 539
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 447 PTEEEEV-VKD-----RTQSKQPQPOPSTGSLTGPPLS-ELSVNSQGVAVANTV 497
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 540 PELTNVQLQAQDYPAPNIMRTQVYRPPPYPPRPNANSTPDLRHLIYISSN---PDLIT 596
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 498 LSPNLSPTDKQASPL-ISPLINDOACPRDDEDEGRKRKRPDTPDKAYFAKEVSTERTYL 556
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 597 RVVHHSVQTFQEDSLPVAHSLEQVSEPLTAARHAHQKR-----635
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 557 KDLEV--ITSWFQSTVSKEDAMPEALKSLFFNFEPFLHKFHTNFKLEIQRALWEGRSN 614
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 636 NSIEIAGLTHGFEGLRKEETM-SASAADVAPRTFSAGSSQSVFSDKVQ-----EG--- 686
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 615 AQIRDYQIRIGDWMKNIQCMKHLAHLWKHSEALE---ALENGIKSSRLNFCRDFELQ 671
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 687 ---TEEQSGSGGYSHKSLSDATMLIHSSEDEDEDDSSREHAVSPRLTAAFSQQQLN 743
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 672 KVCYLPNTFLRLPLRLHMYKQVLERLCKHHPSPHADFRDCRAALAEITEMVAQLHGTM 731
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 744 YPC--ASVTPVTGPLH-----IFP-----PKSHVTEPEKRA-----KDLSPVH--- 779
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 732 IKMENFQ-KLHELKDKLDIGIDNLVWVGREFIRLGLSK-----LSGK-----G 773
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 780 LVMETHQPRRHGLLTPSMSESLLTTSGRYRARRDSLKKRPVSDLLSGKKNTVBGLPLGG 839
      | : | : | : | : | : | : | : | : | : | : | : | : |
QY 774 LOORMEFL--FNDVLLYTSRGLTASNQFVKHGLPOLPYGMTIIESEDEWGVPHCLTLRQ 831
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 840 MKKTRADAKKIGPLKLAALNGLSL-----RLPLPDEGKEVSTRATNDRCKVLEQRL 892
```

QY 832 OSIIVAASSRSEMEKVEDIQMAIDLAESKSSPAPEFLASSPPDNKSPDEATAQDQESD 891  
 Db 893 EQGTVFTFYERILKRLVD-----GECSTARLPE-----NAERNRFDQ 930  
 QY 892 DLSASRSLERQAPHRGNTGVHVCWRHNTSVSMVDFS-IAVENOLSGNLLRKFNKNSGWQ 950  
 Db 931 VLPYDDARVELVPTKENGTGINASHLKVSGLIEWDIATQGPLO-NTCDF-----WQ 984  
 QY 951 KLWVFTNFCFLFFYKSHODNHPPLASLLGYSLTIPSESENIQDYVFKLHFKSHVYFR 1010  
 Db 985 MVW-----EQGVAILAM-----VTAEEGGRE-----KSFYWR 1014  
 QY 1011 AESY-----TFERNMEVIR-----SATSSASRPHVLSHESLVY 1045  
 Db 1015 LGSRHNTVTVYGRFKITTRFTDSCYATTGLKMKHLLTGQERTVM 1059

RESULT 15  
 ID PTNL\_MOUSE STANDARD; PRT; 1176 AA.  
 AC Q62136;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 21 (EC 3.1.3.48)  
 DE (PROTEIN-TYROSINE PHOSPHATASE PTP-RL10).  
 GN PTPN21.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Liver;  
 RX MEDLINE=95140431; PubMed=7838537;  
 RA Higashitsuji H., Arii S., Furutani M., Imamura M., Kaneko Y.,  
 RA Takenawa J., Nakayama H., Fujita J.;  
 RT "Enhanced expression of multiple protein tyrosine phosphatases in the  
 RT regenerating mouse liver: isolation of PTP-RL10, a novel cytoplasmic-  
 RT type phosphatase with sequence homology to cytoskeletal protein  
 RT 4.1.";  
 RL Oncogene 10:407-414(1995).  
 CC -1- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF GROWTH AND  
 CC DIFFERENTIATION OF LIVER CELLS.  
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =  
 CC PROTEIN TYROSINE + PHOSPHATE.  
 CC -1- TISSUE SPECIFICITY: LIVER.  
 CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-  
 CC TYROSINE PHOSPHATASE FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: D37801; BAA07053.1; -  
 CC HSSP: P18031; 1PTT.  
 CC MGD: MGI:134406; Ptpn21.  
 CC InterPro: IPR000299; Band\_4.1.  
 CC InterPro: IPR000387; Tyr\_phosphatase.  
 CC InterPro: IPR000242; Tyr\_prot\_phptase.  
 CC Pfam: PF00373; Band\_41; 1.  
 CC Pfam: PF00102; Y\_phosphatase; 1.  
 CC PRINTS: PR00700; PRTYPHPTASE.  
 CC PRINTS: PR00935; BAND41.  
 CC SMART: SM00295; B41; 1.  
 CC SMART: SM00194; PTPc; 1.  
 CC PROSITE: PS00660; BAND\_41\_1; 1.  
 CC PROSITE: PS00661; BAND\_41\_2; 1.

DR PROSITE; PS50057; BAND\_41\_3; 1.  
 DR PROSITE; PS50057; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
 KW Structural protein; Cytoskeleton; Hydrolase.  
 FT DOMAIN 77 241 BAND 4.1-LIKE  
 FT DOMAIN 923 1176 PROTEIN-TYROSINE PHOSPHATASE.  
 FT ACT\_SITE 1110 1110 BY SIMILARITY.  
 FT DOMAIN 340 343 POLY-PRO.  
 FT DOMAIN 565 572 POLY-PRO.  
 SQ SEQUENCE 1176 AA; 133490 MW; 529FBE22F1335B75 CRC64;

Query Match 7.3%; Score 401.5; DB 1; Length 1176;  
 Best Local Similarity 20.7%; Pred. No. 2.2e-15;  
 Matches 248; Conservative 167; Mismatches 459; Indels 325; Gaps 45;

QY 25 LERGOKPPPTSGKLVSITKIQMLDDTQEAPEVPOAPGKVLDDAVCNHLNVEGDYGLE 84  
 Db 9 LKTRRYTVSKSCLVA-RIQLNNEVFETLSVSGOESLEAVORLEUREVYFSLW 67  
 QY 85 FPDHKKITVMDLLKPIVKQIRRPKHVV---VKFWAKFFPDHDTQLQBELTRYLFALQVK 141  
 Db 68 YNKNQORRWVDLEKPLKKQL--DKHALEPTVYFGLVYFVSQLOQEIIRYQYIQLK 125  
 QY 142 QDLAQGRLTNDTSAALLISHIVSGEIGDDEALDRHLAKNKYIPOODALEKIVE--- 198  
 Db 126 KDILEGNLPTCLEHAIQLAGLAVQDFQDYESODFLQKFAALLPVAWLQDERVLEEA 185  
 QY 199 ----FHHNHIGOTPAESDFOLLEIARRLEMYGIRLHPAKDREGTKINLAVANTGILV-- 251  
 Db 186 QKVALHOKYRGLTAPAEALLYMOEVMGEGEESTPAKDSQGSIDISIGACLDGIFVKH 245  
 QY 252 FQGFTKINAFNNAKVRKLSFKRKLRLKRPDANSAYQDTLEFLMASRDFCKSFKWICVE 311  
 Db 246 KNGRPV-VFRWHDIANMHNKSFFALEL--ANK--EETIQFOTEDMETAKYVRLCVA 299  
 QY 312 HHAFRL-----FEPPKP-----PKVLPFRSGSFRFSRGTQ 344  
 Db 300 RHKPYRLNOCSLQTAATLNSVRDSSRSLPKPQPPQYAMPPPQLHYNGHYTEPPASSQ 359  
 QY 345 QVLDYVKEG-GHKVKQFER-----KHSKHSIRS-----LASOP 378  
 Db 360 DNIFVPKNKGFYCHSQTSLDRTQIDLSGRINGSVYSAHSTNSLNTLOPYLPSPMSNP 419  
 QY 379 TELNSEVLE---SQSQSTSLTFEGEAESGGQS---CRRG----- 412  
 Db 420 SITGSDVMRPDLSFHRHSALIPSYRPTPDYETVMKQLNKGVMHADRHSHSLRLNIGS 479  
 QY 413 -----KEPKVSAGEPGSH-----SPAPRRSPAGNKQADGAASA 446  
 Db 480 SYAYSRPDALVYSQPEIREHPHLTSPQSAHYFPFNLYSFHSQSPYPYPAERPPVGVSV 539  
 QY 447 PTEEEEV-VKD-----RTQSKPPOPQPGSTGLTSPHLS-ELSVNSGGVAPANVT 497  
 Db 540 PELTNVQLQADYPAPNIMRQVYRPPPPYPPRANSTPDLRHLIYSSNN--PDLIT 596  
 QY 498 LSPNLSPTDKQASPL-ISPLNDQACPRTDDEDEGRKRKRPDTPKAYFIAKEVSTERTYL 556  
 Db 597 RRVHHSYQTFQEDSLPVAHSLQEVSEPLTAARHAHLQKR-----NSIETAGLTHGFEGLRL 652  
 QY 557 KLEVITVWFQSTVSKEDAMPEALKSLIFPNFELHKEHFNFLKEIQRLALWEGRSNAQ 616  
 Db 653 KERTV-----SASAADVAPRTFSA-----GSSSV 677  
 QY 617 IRDYQRIGDVMKNIQGMKHLAAHLWKHSEALEALENGIKSRLENFCRDF----- 668  
 Db 678 FSDKMK-----QEGTEQEGGRYSHKXSLSDATMLIDSEDEDELEEDSSREQAISA 729  
 QY 669 -----ELQKVCYLPLNTFLRLPLRLHMLHYKQVLERLCKHHPPSHADPRCRAA 716  
 Db 730 VSEPLRTAAFSQELNYPG--ASATPITGPLH-----IFE-----PKPHVTEPRRAK 774





GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:49:58 ; Search time 77.11 Seconds  
(without alignments)  
1982.293 Million cell updates/sec

Title: US-09-555-342A-2  
Perfect score: 5463  
Sequence: 1 MGIEORPPGSLGAPENS.....SATSSASRPHVLSHKESLVY 1045

Scoring table:  
BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL.17.\*

- 1: sp.archaea.\*
- 2: sp.bacteria.\*
- 3: sp.fungi.\*
- 4: sp.human.\*
- 5: sp.invertebrate.\*
- 6: sp.mammal.\*
- 7: sp.mhc.\*
- 8: sp.organelle.\*
- 9: sp.phage.\*
- 10: sp.plant.\*
- 11: sp.rodent.\*
- 12: sp.virus.\*
- 13: sp.vertibrate.\*
- 14: sp.unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	5463	100.0	Q9Y4F1	Q9Y4f1 homo sapien
2	2941.5	53.8	Q94887	Q94887 homo sapien
3	2117	38.8	Q99KU3	Q99ku3 mus musculus
4	913.5	16.7	Q9VN69	Q9vn69 drosophila
5	867	15.9	Q9VN68	Q9vn68 drosophila
6	850.5	15.6	Q17905	O17905 caenorhabdi
7	746.5	13.7	Q9WV92	Q9wv92 mus musculus
8	746.5	13.7	Q9JMB2	Q9jmb2 rattus norv
9	743.5	13.6	Q9BRP5	Q9brp5 homo sapien
10	735.5	13.5	Q95713	Q95713 homo sapien
11	732	13.4	Q970318	Q970318 mus musculus
12	728.5	13.3	Q9Y2J2	Q9y2j2 homo sapien
13	728	13.3	Q9JMB3	Q9jmb3 rattus norv
14	720	13.2	Q43491	Q43491 homo sapien
15	711.5	13.0	Q9R102	Q9r102 mus musculus
16	708.5	13.0	Q9Y579	Q9y579 homo sapien
17	704	12.9	Q9N179	Q9n179 bos taurus
18	700	12.8	Q9WTP1	Q9wtp1 rattus norv
19	699	12.8	Q9H4G0	Q9h4g0 homo sapien

20	698.5	12.8	641	4	Q14245	Q14245 homo sapien
21	695.5	12.7	879	11	Q9Z2H5	Q9z2h5 mus musculus
22	694	12.7	934	4	O15046	O15046 homo sapien
23	692.5	12.7	879	11	Q9WTF0	Q9wtf0 rattus norv
24	691.5	12.7	926	11	Q9WU22	Q9wu22 mus musculus
25	691.5	12.7	1698	5	Q24440	Q24440 drosophila
26	691.5	12.7	1698	5	Q9V8R9	Q9v8r9 drosophila
27	686.5	12.6	230	4	Q9UFE7	Q9ufe7 homo sapien
28	685	12.5	703	5	Q9V8S0	Q9v8s0 drosophila
29	682	12.5	889	5	Q9V8R8	Q9v8r8 drosophila
30	626	11.5	687	4	Q9H975	Q9h975 homo sapien
31	609.5	11.2	505	6	Q9MYU8	Q9myu8 canis famli
32	608	11.1	504	11	Q99KZ8	Q99kz8 mus musculus
33	586.5	10.7	972	5	Q9VFU8	Q9vfu8 drosophila
34	581	10.6	913	4	Q9H329	Q9h329 homo sapien
35	568	10.4	527	11	Q9JMC8	Q9jmc8 mus musculus
36	568	10.4	568	5	Q9GP93	Q9gp93 caenorhabdi
37	562	10.3	504	4	Q9H328	Q9h328 homo sapien
38	562	10.3	518	4	Q9P2V3	Q9p2v3 homo sapien
39	561	10.3	987	5	Q9NDP4	Q9ndp4 clona intes
40	534	9.8	440	4	Q9NX84	Q9nx84 homo sapien
41	524	9.6	283	4	Q9BQV2	Q9bqv2 homo sapien
42	520.5	9.5	766	11	O88387	O88387 rattus norv
43	518.5	9.5	598	4	Q9HCS5	Q9hcs5 homo sapien
44	465.5	8.5	414	5	Q20653	Q20653 caenorhabdi
45	460.5	8.4	699	4	Q9P215	Q9p215 homo sapien

## ALIGNMENTS

RESULT 1

Q9Y4F1 PRELIMINARY; PRT; 1045 AA.  
ID Q9Y4F1  
AC Q9Y4F1;  
DT 01-NOV-1999 (TRENBLrel. 12, Created)  
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
DE CDEP.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=CARTILAGE;  
RX MEDLINE=98086358; PubMed=9425278;  
RA Koyano Y., Kawamoto T., Shen M., Yan W., Noshiro M., Fujii K.,  
RA Kato Y.;

RT "Molecular cloning and characterization of CDEP, a novel human protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of Rho guanine nucleotide exchange factors.";  
RL Biochem. Biophys. Res. Commun. 241:369-375(1997).

DR EMBL; AB008430; BAA24267.1;  
DR HSSP; P08567; IPLUS.  
DR InterPro; IPR000299; Band\_4.1.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR000219; RhoGEF.  
DR Pfam; PF00373; Band\_41; 1.  
DR Pfam; PF00169; PH; 2.  
DR Pfam; PF00621; RhoGEF; 1.  
DR PRINTS; PR00935; BAND41.  
DR SMART; SM00295; B41; 1.  
DR SMART; SM00233; PH; 2.  
DR SMART; SM00325; RhoGEF; 1.  
DR PROSITE; PS00660; BAND\_41\_1; UNKNOWN\_1.  
DR PROSITE; PS50057; BAND\_41\_3; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 2.  
SQ SEQUENCE 1045 AA; 118632 MW; OE8B2D61C0F58417 CRC64;

Query Match 100.0%; Score 5463; DB 4; Length 1045;  
Best Local Similarity 100.0%; Pred. No. 0;



```
Db 1 MGEIEGTVRVLOTAGMRLGAQTPGVSTLEPCQTLPRMQEKHLHLRVKLLDNTMEIFDI 60
Qy 57 QORAPGKVLDAVCHNLNLVGDYGLGEBPDHKKITVWLDLLKPIVKQIRPKHVYVVF 116
Db 61 EPKCDQVLLTQVWRKRLNVECDYGEFQNTQSWIWEPMKPIQIRRPKNVRLA 120
Qy 117 VKFFPDHTQOEELTRVLFALQVQDLAOGRLTCNDTSAALLISHIVQSEIGDDEALD 176
Db 121 VKFFPDGQLOEETRYLFALQKRLDLEERLTACADTAALLTSHLQSEIGDYDETL 180
Qy 177 REHLAKNYIPOQDALEDKIVFEHFNHIGQTPAESDFOLLEIARLEMYGIRLHPAKDRE 236
Db 181 REHLAKNEYLPQOCHLEKILEFHQKHVGQTPAESDFQVLEIARLEMYGIRFHASDRE 240
Qy 237 GTKINLAVANTGILVFGFTKINAFNAWKVKRLSKRRLIKLRDPDANSAYQDTLEFLM 296
Db 241 GTKIQLAVSHMGVLFQGTTKINTENWSKVRKLSFKRRLIKLHPVHGVPYQDLEFL 300
Qy 297 ASRDFCKSPWKTVEHHAFFRLEFPKPKPVLSRGSSEFSGRTOKQVLDYVKEGGH 356
Db 301 GSRDECKNFKVICVEYHTFFRLDDQPKPAKAVFFSRGSSFRYSRTOKQVLDYFKDSGM 360
Qy 357 KKVQPERKHSKIH-SIRSI-ASQPTLELNVLEQSQSTSLTFGGAESPQSGQSCRR--- 411
Db 361 KRIPYERRSKTHTSVRALTLDPKQ-----SISFPEGLRTPASPANAFY 407
Qy 412 --GKEPKVSAGBPG-----SHPSPAPRRSPAGNKQADGAASAPTEEBEEVVKDRTOQ 461
Db 408 SLSPSTLVPGLPEFKDSSSLTDQVSVKSPAAERKSGAVAGGPD-----TPS 457
Qy 462 SKPQPP---QPTSGSLTSGPSHLSVNSQGVAVANTVLSNLSNLPDTPKQASPLISPLLN 518
Db 458 AOPGLPAPALQPGGLSTKSPQSPSSRKSPLSLSPA---FQVPLGPAEGSGSPLSPVL 514
Qy 519 DOACPTDDEGRKRKPTDKAYIAKEVSTERTYTKLDEIVTSWFOSTVSKEDAMPE 578
Db 515 DAGGAGMCE-EPRKRVPADEAYIIVEILATERTYTKLDEIVTSWFAVVKEDAMPA 573
Qy 579 ALKSLIFNPEFLKPHNTFLKEIQRLALWEGRNAQIR-DYQRIGDVMKNIQMKHL 637
Db 574 TLMWTLFSNIDPIYEFHGFLEVEORLALWEGPSKAHTKGSHQIRIGDILLRNMQLEF 633
Qy 638 AALWKHSEALEALENGIKSSRLENFCDLFEQKVCYLPNTFTLRLHLRMLHKVQVLE 697
Db 634 TSYFORHDEVLELEKATKRCKKLEAVYKEFELQKVCYLPNTFTLKLQIRLLHYRLLR 693
Qy 698 RLCKHHPHADFRDCRAALAEITMVAQLHCTMKMENFQKLHELKLDLIGDNLVVP 757
Db 694 RLCGHYSGHHYADCHDAKATEVTTLQHLIRLENLQKLTELQRLDVLGIENLIAPG 753
Qy 758 REFIRGLSKLSKGLQRMFFLNDVLLYTSRGLTASNQKFKVHQPLYGMTIEESD 817
Db 754 REFIREGLHKLTKGLQRMFFLSDMLLYTSKGVAGTSHFRIGLLPLQGLMVEESDN 813
Qy 818 EGVPHCLTLRGQROSIIVAAASRSEMEKWEVDIOMADLAEKSSPAPEF---LASSP 873
Db 814 EWSVPHCFETIYAORTIIVAASTRLEKEKMDLNSAIOAAKSGGDAPALPGRTVCTRP 873
Qy 874 PDKSPDEATAADQESDDLSASRTSLERQAPHRGNTMVHVCWHRTTSVSMVDEFAVEN 933
Db 874 P--RSPNE-VSLQESEDARCVRSLEGGHQRANTTHVCWYRNTSVSRADHSAAVEN 930
Qy 934 QLSGNLLRKFKNKSNQKLVWVFTNFCFFYKSHQDNHPLASLPLLGYSLTIPSESENIQ 993
Db 931 QLSGYLLRKFKNKSNQKLVWVFTNFCFFYKTHQDDYPLASLPLLGYSVSIPREADGLH 990
Qy 994 KDYVFKLHFKSHVYFRAESETTFRMWEVIRSATSSASR 1033
Db 991 KDYVFKLQFKSHVYFRAESKYTFRMWEVIRSATSSASR 1030
```

RESULT 3  
Q99KU3

```
ID Q99KU3 PRELIMINARY; PRT; 420 AA.
AC Q99KU3;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE UNKNOWN (PROTEIN FOR IMAGE:3493093) (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004009; AA04009.1; -.
FT NON_TER 1
SQ SEQUENCE 420 AA; 48387 MW; 804411A6B6613F96 CRC64;

Query Match 38.8%; Score 2117; DB 11; Length 420;
Best Local Similarity 95.2%; Pred. No. 1.4e-144;
Matches 400; Conservative 10; Mismatches 10; Indels 0; Gaps 0;

Qy 626 VMLKNIQGMKHLAAHLWKHSEALEALENGIKSSRLENFCDLFEQKVCYLPNTFTLRLP 685
Db 1 VMLKNIQGMKHLAAHLWKHSEALEALETSIKSRLEHFCRDFELQKVCYLPNTFTLRLP 60
Qy 686 LHRLMHYKQVLERLCKHHPPSHADFRDCRAALAEITMVAQLHGTMIKMFQKLHELK 745
Db 61 LHRLMHYKVLRLCKHHPPNHADFRDCRAALAEITMVAQLHGTMIKMFQKLHELK 120
Qy 746 DLIGDNLVVPQREFIRGLSKLSKGLQRMFFLNDVLLYTSRGLTASNQFKVHGOL 805
Db 121 DLIGDNLVVPQREFIRGLSKLSKGLQRMFFLNDVLLYTSRGLTASNQFKVHGOL 180
Qy 806 PLYGMTIEESEDGVPCHLTLRGQROSIIVAAASRSEMEKWEVDIOMADLAEKSSPA 865
Db 181 PLYGMTIEESEDGVPCHLTLRGQROSIIVAAASRSEMEKWEVDIOMADLAEKSGPT 240
Qy 866 PELFASSPDNKSPEATAADQESDDLSASRTSLERQAPHRGNTMVHVCWHRTTSVSMV 300
Db 241 PELFASSPDNKSPEATAADQESDDLSASRTSLERQAPHRGNTMVHVCWHRTTSVSMV 300
Qy 926 DFSIAVENOLSNLRKFKNSGWOKLVWVFTNFCFFYKSHQDNHPLASLPLLGYSLT 985
Db 301 DFSIAVENOLSNLRKFKNSGWOKLVWVFTNFCFFYKSHQDNHPLASLPLLGYSLT 360
Qy 986 PSEENIKDYYFKLHFKSHVYFRAESETTFRMWEVIRSATSSASRPHVLSHESLVY 1045
Db 361 PSEENIKDYYFKLHFKSHVYFRAESETTFRMWEVIRSATSSASRAHLSHESHLY 420

RESULT 4
Q9VNE9 PRELIMINARY; PRT; 562 AA.
AC Q9VNE9;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE CG2008 PROTEIN.
GN CG2008.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
```



RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*,"  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003604; AAF52081.1;  
DR Flybase: FBgn0037286; CG1283.  
DR InterPro: IPR000299; Band\_4.1.  
DR Pfam: PF00373; Band\_41; 1.  
DR PRINTS: PR00935; BAND41.  
DR SMART: SM00295; B41.1;  
DR PROSITE: PS00660; BAND\_41\_1; 1.  
DR PROSITE: PS00507; BAND\_41\_3; 1.  
SQ SEQUENCE 330 AA; 38707 MW; FBDABF30BC86CBF CRC64;

Query Match 15.9%; Score 867; DB 5; Length 330;  
Best Local Similarity 51.5%; Pred. No. 1.5e-54;  
Matches 168; Conservative 64; Mismatches 84; Indels 10; Gaps 6;

QY 46 MLDLDTQAEFVPPQAPGKVLVDVAVCNHNLNVEGDYFGLFEPD---HKITVWLDLLKPIV 102  
DB 1 MLDSDITWFQVQAKALGRVLFQVCRQNLNLEADYFGLYQEVSTHK--YWLDLEKPMN 58  
QY 103 KQIRPK-HVVVKYVKKFPDPDHTLOELTRYLFALQVKQDLAQRLTCDNTSAALLIS 161  
DB 59 RQVGLSLDPLVRFICFVTPDPAOLEEYTRYLFCLQIKRDLATGSLQCNDAALMAS 118  
QY 162 HIVQSEIGDF--DEALDREHLAKNKYIQQDQ--LEDKIVEFHNNHIGOTPAESDQLLEI 218  
DB 119 YIVQACSGDFDPEDYDHTYLSYRFVNPQDQATMORKIMENHKHVGQSPAEADNLLET 178  
QY 219 ARRLMEYGLRLPAKDRGKTLNLAVALTILVFOGETKINAFNWKVSKLSEKPKRELI 278  
DB 179 ARRCELYGKMKHPAKDVEGVPLNLAVHMGITVFNITRINTFSWAKRKISFKRKLIV 238  
QY 279 KLRPDANSAYQDTLEFLMASRDFCKSFWKICVEHHAFFLEF-EPKPKPKVLFVSRGSSF 337  
DB 239 KLUHPEGYGYKDTVEFFEGNECKNFKKVENHGFFRCTAVQNTPRKTRVLSRGSF 298  
QY 338 RSGRTQKQVLDYVKEGGHKVQFER 363  
DB 299 RYSGTKQKQIEFVRENYVKRQNFQ 324

RESULT 6  
ID O17905 PRELIMINARY; PRT; 1111 AA.  
AC O17905;  
DT 01-JAN-1998 (TRENBLrel. 05, Created)  
DT 01-JAN-1998 (TRENBLrel. 05, Last sequence update)  
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)  
DE H05G16.1 PROTEIN.  
GN H05G16.1.  
OS *Caenorhabditis elegans*.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoides;  
OC Rhabditidae; Peloderinae; *Caenorhabditis*.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA White S.;  
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94150718; PubMed=7906398;  
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,  
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,  
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
RA Smalton N., Smith A., Sonhammer E., Staden K., Sulston J.,  
RA Thierley-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,  
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans*,"  
RL Nature 368:32-38(1994).  
DR EMBL: Z97190; CAB10024.1; -.  
DR InterPro: IPR000299; Band\_4.1.  
DR Pfam: PF00373; Band\_41; 1.  
DR PRINTS: PR00935; BAND41.  
DR PROSITE: PS00660; BAND\_41\_1; UNKNOWN\_1.  
DR PROSITE: PS00507; BAND\_41\_3; 1.  
DR SMART: SM00295; B41; 1.  
SQ SEQUENCE 1111 AA; 124808 MW; 1EE53EB512C0FE2A CRC64;

Query Match 15.6%; Score 850.5; DB 5; Length 1111;  
Best Local Similarity 26.6%; Pred. No. 1.3e-52;  
Matches 287; Conservative 157; Mismatches 363; Indels 271; Gaps 38;

QY 22 ISTLERGQKPPPTPS-----GKLVSIKIQMLDDTQEAPE----- 55  
DB 1 MSNIPRGVCAPPPCCMNSKRGRLVCIKVRLDDTVAVPHLGVICFFCVSVSEHGTYYFL 60  
QY 56 VPQAPAGKVLVDVAVCNHNLNVEGDYFGLFEPDHPKKITVWLDLLKPIVKOIRR-PKHVVVK 114  
DB 61 LQHKAIQGTLLDEVCRLNLLCEDYFGLSFIDINGNHCWLDREKTLRLQITNGSTDAFY 120  
QY 115 FVVKFPDPDHTLOELTRYLFALQVKQDLAQRLTCDNTSAALLISHIVQSEIGDF--D 172  
DB 121 FVVKFYTNPIDLEEYTRYLFTMQIKRDLALGEUHCSDNTASLLSAYLVSECDFSE 180  
QY 173 EALDREHLAKNKYIQQDQ--ALEDKIVEFHNNHIGOTPAESDQLLEIARREMYGIRLHP 231  
DB 181 DYPDATYLSHTRFVNPQTLFQKVMNDHNRNFIGMTGESDLAMLEVARCCDFYGVKLAH 240  
QY 232 AKDRGKTLNLAVALTILVFOGETKINAFNWKVSKLSEKPKRELIKLRPDANSAYQDT 291  
DB 241 AKDIDGNDAAALSVMLGKIVFRQLQDITFWSWARTKLSFKRKLKLLVLPDYSYQLKET 300  
QY 292 LEFLMASRDFCKSFWKICVEHHAFFLEFEEPKPKPKVLF-SRGSSEFRFSGRTQKQVLDY 350  
DB 301 VEFSETRDECKNFKWKICVEHHAFFRCVQAEPEPKRETRFFISKGSFRHGTQQLIDY 360  
QY 351 VKEGGHKVQFER-----KHSKI 368  
DB 361 VREHKRREPTRPLRSAASTRKGTYSSTYGLVDRPTKHRNGSVYEPNQTDYKNHQT 420  
QY 369 HS-----IRSLASQPE-----LNSEVLEQ---SQOSTSLTFGEAGSPGGSCRRG---- 412  
DB 421 HSSMPHIAHISQPADHSFGTLDARVGKSVTRDPSTSVTQRLRLKRSERCISDVDTV 480  
QY 413 KEPKVSAGEPGSHSPAPRRSPAGNKQADGAASAPT--EEEEVVKDRTOQSKPOPPQPS 470  
DB 481 ERQPSRCRPVSHAINI--SSTSSKEQPLSVSLPNVLSDDLQMVCKEIEIQNDPPKSV 538  
QY 471 TG-----SLTGPSPHISELSVNSQGGVAP--ANVTLSPNLSPDTQKAS 510  
DB 539 SGDNFQQRSSRDYDNVSEDSYRLSDH--ERSTRSEVGVGSKFAAATFNSFTVARKPCS 596  
QY 511 PLISPLL-NDQACPTDDEDEGRK-----RFTDKAYFAKE----- 547  
DB 597 NVVKRWVAHTKSTPNSTDDEGALKSASEYQTFRIKEYFPFARNANIVPIEIDGPNVDLS 656  
QY 548 -----VSTTERTYLKDEVITSWFQSVSKEDAMPALKSLIFPN-PEPLHKHETNLF 599  
DB 657 ARRSPPAATTTTRTIVANTTTTTSKVLTSAGAVL---MKPKVISNDHEPSSGTHSS--- 711  
QY 600 KEIEQRLALWEGRSNAQRDYORIGDV-----MLKNIQGMKHLAAH----- 640  
DB 712 -----RVSPESCTYCALGPLQVRVITKENMMITPEGFEKKAKPNPPKPSFP 759

```
QY 641 -LWKHSEALEALENGKSSRRLENFCRDFELQKVCYLPNTFLRLPHRLMHYKQVLERL 699
Db 760 VQAFHSETEVRE--IKTERASS-----HPLVHMOE--ETP 791
QY 700 CKHPPSHADFRCRAALAEITEWVAQLHCTMIKMFNQKLHELKDLIDGLNVVPGRE 759
Db 792 FSRAPLKFDKCKRSLSRALISVQSDN----PDVQKCHLFSND-----IP--- 836
QY 760 FIRGLSLKSGKGLQORMFLENFLLVTSRG-----LTASQKFKVHGOLPL--- 807
Db 837 -----YILTMRNVENTOSLPYSSFKDVSTAKKQGYESNLKRVSKSPFRRKSLDLVPRK 892
QY 808 -----YGMTIESEDEWGPCHL-----TLRQROSIIVAASSRSEMEKMWEDIDM 853
Db 893 RLPSPGNFSAQDHTISPTTPDSVLEVLLRRRLGDK-SVIKTKSKRTDPRRQTPVR- 950
QY 854 AIDIAKSSSPAPEFLASSPP-----DNKSPDEATAQDESDLLSASRTSLERQAPH 906
Db 951 -FDLPSPSCSPT--AGGSTPFTISILNDLDFECVSESRLHEDMD-----RLDKTTTPH 1000

RESULT 7
Q9WV92
ID Q9WV92 PRELIMINARY; PRT; 929 AA.
AC Q9WV92;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROTEIN 4.1B.
GN EPB4.1L3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Parra M.K., Chan N.L., Gascard P.D., Walensky L.D., Lee G.L.,
RA Chasis J.A., Snyder S.H., Narla M., Conboy J.G.;
RT "Characterization of protein 4.1B, a new gene in the protein 4.1
RT family with high level, focal expression in brain.";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF152247; AAD38048.1; -
DR MGD; MGI:103008; Epb4.113.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS50057; BAND_41_3; 1.
SQ SEQUENCE 929 AA; 103337 MW; F4975FF405DA44AE CRC64;

Query Match 13.7%; Score 746.5; DB 11; Length 929;
Best Local Similarity 32.6%; Pred. No. 3.3e-45;
Matches 205; Conservative 100; Mismatches 230; Indels 93; Gaps 22;

QY 18 ENSGISTLERGQ----KPPPTPSGKLVSKIOMLDDTQAEFVQRPAGKVLDDAVCNHL 73
Db 97 QRSSSSKLSRSLPKIVKRP-----KSMQCKVTLLDGEYGCDDVKRSRGQVLFDKVCEHL 151
QY 74 NLVEGDFYFGLFPPDHKKITVWLLDKLPVKQIRRPK-HVVVKFVVKFFPPDHTQLOEELT 132
Db 152 NLEKDYFGLTYRDAENQKNWLDPAKEIKQIRSGAWH--FSFNVKFYPPDPAPQLSEDT 209
QY 133 RYLFALQVKODLAQGRLTCDNTSAAALLSHIVQSEIGDFD-EALDRDLAKNKNIPQ-Q-D 190
Db 210 RYLYLQLRDRDIVSGRLPCSFVTVLALLGSYTVQSLDGYDPDCGNDYISEFRFAPNHTK 269
QY 191 ALEDKIVFHHNIGQTPAESDFQLLEIARRLEMYGIRLHPAKDRGTKLNLAVENTGIL 250
Db 270 ELEDKVIELHKSIRGMTPAEAMHLENAKKLSMYGVDLHAKDSEGVETMLGVCSGLL 329
```

```
QY 251 VFQGTKIINAFNWKVKLSFKRRKRELKLRDPDANSAYQDTLEFLMASRDFCKSFWKICV 310
Db 330 IYRDLRLINRFAMPVKLKISYKRNNFYIKIRPGEPFQFESTIGFKLPNHRRAKRLWKVCV 389
QY 311 EHHAFRLFEKPKPKPVLFSGSSFRSGRTQKQVLDYVKEGGHKKVQFERKKSHIHS 370
Db 390 EHHTEFRL-LPEAPPKFL-TLGSKFRYSGRTQAQTRASALIDRPAPYFEFSSSKRYT 447
QY 371 I-RSLASQTELNSEV-LEQSOOSTSLTSGEAGSPGGQSC-----RRGKEPKVSA 419
Db 448 MRSLDGASVENHEIYMKDSVSAEVTGQYATTIGISQTNLITVTPEKKAEEERVEE 507
QY 420 GE-----PGSHPSAPRR-----SPAGNKQADGAASAPTEEEVEVVK-----DRTQOS- 462
Db 508 EDRKKAEEATPVTLRHGKTDSERTDAADGETSATESDQEDAEIKAAQDLDKTQDEL 567
QY 463 -KPQPPQSTGSLTSPHLSSELSVNSQGVAPAN-----VTLSPNLSPDTKQASPLISPL 516
Db 568 MKHQ-----TNISELKRFTLETSTALTNEWEKRLSTSPVRLAARQEDAPMIEPL 618
QY 517 LNDQACPRTDDEDEGRR-----KRPETKAYFIAKEVSTTERTYTKDLEVIT 563
Db 619 V-----PEETKQSGGKLMGDSHLSLESARKPTE---FIGVVSSTQ----- 659
QY 564 SWFQSTVSKEDAMPEALKSLIFNFEPL 591
Db 660 SWVQKLETKTPEVPEAEVEST--PHQPPL 685

RESULT 8
Q9JMB2
ID Q9JMB2 PRELIMINARY; PRT; 963 AA.
AC Q9JMB2;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TYPE II BRAIN 4.1 MINOR ISOFORM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE=20267856; PubMed=10806359;
RA Yamakawa H., Ohara O.;
RT "Comparison of mRNA and protein levels of four members of the protein
RT 4.1 family: The type II brain 4.1/4.1B/KIAA0987 is the most
RT predominant member of the protein 4.1 family in rat brain.";
RL Gene 248:137-145(2000).
DR EMBL; AB032828; BAA90775.1; -
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS50057; BAND_41_3; 1.
SQ SEQUENCE 963 AA; 107071 MW; B37C137B04305F71 CRC64;

Query Match 13.7%; Score 746.5; DB 11; Length 963;
Best Local Similarity 33.7%; Pred. No. 3.5e-45;
Matches 198; Conservative 100; Mismatches 222; Indels 67; Gaps 18;

QY 18 ENSGISTLERGQ----KPPPTPSGKLVSKIOMLDDTQAEFVQRPAGKVLDDAVCNHL 73
Db 92 QKSSSSKLSRSLPKIVKRP-----KSMQCKVTLLDGEYGCDDVKRSRGQVLFDKVCEHL 146
QY 74 NLVEGDFYFGLFPPDHKKITVWLLDKLPVKQIRRPK-HVVVKFVVKFFPPDHTQLOEELT 132
Db 147 NLEKDYFGLTYRDAENQKNWLDPAKEIKQIRSGAWH--FSFNVKFYPPDPAPQLSEDT 204
```

Qy	18	ENGISTLERG-----QKPPPTSGKLVSTIKOMLDDTQEAPEVQAPGKVKLLDVA	CNH	72
Db	89	QKSSSKLSRPLKIVKKP-----KSMQCKVILLDGSSEVTCDEVKSRGQVLF	DKVCEH	142
Qy	73	LNIVEGDYFGLRPPDHKKITVWLDDLKPIVKQIRPK-HVVYKVVVYKFPDPDHTQLQ	QEEL	131
Db	143	LNLEKDYFGLTVRDAENQKNWLDPAKETKKQVRSAAWH--FSFNKVFYPPDPAQLS	SEDI	200
Qy	132	TRYLFALQVQKDLAQGRKLCNDTSAALLSHIVQSEIGDFD-EALDREHLAKNKYIPQ	O-	189
Db	201	TRYVLCQLRDDIVSGRLPCSFVTLALLGSYTVQSELDGVDYDPCDGSVDYISEFR	PANHT	260
Qy	190	DALEDKIVFHHNHIGOTPAESDFOLLETARRLEMTGIRLHPKADREGTKINLAVANT	GCI	249

[illegible]



```
Db 122 YTVQSEGLDYPDECGSDYISEFRFAPNHTKLEDEKVIELHKGSRGTMTPAEAEHMFLENA 181
QY 220 RRLMEYIIRLHPAKDRGCTINLAVANTGILVFGQFTKINAFNNAWKVRLSEFKRKRFLIK 279
Db 182 KKLISMYGVDLHHAKDSGEVIMGVCASGLIIVDRURINRFPWPKVKLISYARNFYIK 241
QY 280 LRPDANSAYODTLEFLMASRDFCKSFWKICVEHHAFFRLEFEEKPKPKPVLFVSRGSSFRF 339
Db 242 IRPGEFFQFESTIGFKLPNRAAKRLWKVCVEHHTFRL-LPEAPPKFL-TLGSKFY 299
QY 340 SGRTOQVLYVEGGHKKVQFERKHSKIHSI-RSLASQTEINSEV-LQSQOSTSLTF 397
Db 300 SGRTOAQTRASALIDRPAPYFERSSSKRYTMSRSLDGASVNHETIYMDSMSAAEVGT 359
QY 398 GEAESPG-----GOSRRCKGKPKVSAGPGSHPSAPRSPAGNKOAG 442
Db 360 GOYATTKIGISQTNLITVTPEKKAEEERDEEDKRRKGEEVTPISAIQHGKTDSEDT 419
QY 443 AASAPTEEEEVVKDRTOQSKPOPPQSTGLTSGPHLSVNSQGVAPAN-----VT 497
Db 420 AADGETTATELEKTQDILMKHQ-----TNISELKRTFLETSTDTTAVTNEWKRLS 470
QY 498 LSPNLSPTQKASPLISPLINDQACPTDDEG 531
Db 471 TSPVRLAARQEDAPMIEPLVPEEKM-ETKTESSG 503

RESULT 11
O70318 PRELIMINARY; PRT; 988 AA.
AC O70318
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROTEIN 4.1G (FRAGMENT).
GN EPB4.1L2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98198473; PubMed=9531554;
RA Walensky L.D., Gascard P., Fields M.E., Blackshaw S., Conway J.G.,
RA Mohandas N., Snyder S.H.;
RT "The 13-kD FK506 binding protein, FKBP13, interacts with a novel
RT homologue of the erythrocyte membrane cytoskeletal protein 4.1.";
RL J. Cell Biol. 141:143-153(1998).
DR EMBL; AF044312; AAC40083.1; -.
DR MGD; MGI:103009; Epb4.1L2.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; UNKNOWN_1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00657; BAND_41_3; 1.
FT NON_TER 988
SQ SEQUENCE 988 AA; 109833 MW; B0367A16C5A2EC05 CRC64;

Query Match 13.4%; Score 732; DB 11; Length 988;
Best Local Similarity 32.7%; Pred. No. 4e-44;
Matches 198; Conservative 91; Mismatches 229; Indels 88; Gaps 18;

QY 34 TPSEGLVSTKIOMLDDTOEAFVFPQAPGVKVLVDVNCNHLNVEGVYFGLFPDHPKKITV 93
Db 205 TKTKTKVLAKVTLTLDGTEYSCDLEKRAKGVLPDRVCEHNLNLEKDYFGLLFODHPQKN 264
QY 94 WLDPALPIYKQIRRPKRVVVKVFPDPDHTQLOQLRYLQVLAQVQDLAGRLTCND 153
Db 265 WLDPAKEIKRQLKNLPLWFT-FNVKYPDPDSQLTEDITRYFLCLQLQRODIASGRLPSCF 323
```

```
QY 154 TSAALLISHIVQSEIGDFD-EALDREHLAKNKYIP-QQDALEDKIVEFHNNHIGOTPAES 211
Db 324 VTHALLGSYTLQAEHGDYDPEYDSIDLQDFQFAPAHKTELEEKSELHKKTRGLSPAQA 383
QY 212 DFOLLEIARLEMYGIRLHPAKDRGCTINLAVANTGILVFGQFTKINAFNNAWKVRLSEFKR 271
Db 384 DSQFLENAKRLSMYGVLDLHHAKDSGEVIMGVCASGLIIVDRURINRFPWPKVKLISY 443
QY 272 KRKRELILKRPDANSAYODTLEFLMASRDFCKSFWKICVEHHAFFRLEFEEKPKPKPVLF 331
Db 444 KRGNFYIKVPAELEQFESTIGFKLPNRAAKRLWKVCVEHHTFYRLV-SPEQPPKTKFL 502
QY 332 SRGSSFRFGSRTQKQVLDVYVEGGHKKVQFERKHSKIHSI-RSLASQTEINSEVLEQSOQ 391
Db 503 TLGSKFYRSRTQAOQFTEASTLIDRPAPQFERASSKRV-RSLDGAPI-----GVQDSPP 557
QY 392 STSLTTEGGA-ESPGQSGRRGKPKVSAGPGSHPSAPRSPAGNKOAG-----442
Db 558 -----GEGSVPGPGVISYTTIQDGRDSSKSP-TKATPLPAEGKKNLTVRVDGNIYVRS 610
QY 443 -----AASAPTEEEEVVKDRTOQSKPOPPQSTGLS 474
Db 611 NLMLEDLRAQEAAILKHQASISELKNFNASTPEPRPSEWEKRRV---TLPPLQPOASS- 666
QY 475 TGSPHLSVNSQGVV--PANVT-----LSPNLS-DTKQASPLI---SPLINDQA 521
Db 667 --HETLVNVEEKKRGVKGKDESVITEEMNGKEMSGPGGPGETRKVEPVAHKDSTLSSES 724
QY 522 CPRTDDEGRRKRFPTDKAYFAKEVSTTERTYTKDLEVITSWFQ-----STVSKED 574
Db 725 SSSSESEEDVGEYQPHHR-----VTEGTIRREEQECDELEBEEPQGGAKVVEREA 775
QY 575 AMPEAL 580
Db 776 AVPDV 781

RESULT 12
QY2J2 PRELIMINARY; PRT; 1115 AA.
ID QY2J2
AC QY2J2;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE KIAA0987 PROTEIN (FRAGMENT).
GN KIAA0987.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 6:63-70(1999).
DR EMBL; AB023204; BAA76831.1; -.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
DR PRINTS; PR00935; BAND41.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS00657; BAND_41_3; 1.
DR SMART; SM00295; B41; 1.
FT NON_TER 1115
SQ SEQUENCE 1115 AA; 123704 MW; E38E3A31E028D4A2 CRC64;
```



```
Query Match      13.3%; Score 728.5; DB 4; Length 1115;
Best Local Similarity 32.1%; Pred. No. 8.6e-44;
Matches 196; Conservative 92; Mismatches 217; Indels 105; Gaps 19;

QY 18 ENSGISTLBERG-----OKPPPTPSGKLVSIKIQMLDDTQEAEPVQAPGKVLDDAVCNH 72
DB 117 QSSSSKLSRSLKIVKRP-----KSMQCKVILLDGSYTCDEKRSQGVLFQVCKVCEH 170

QY 73 LNLVEGDYFLEFPDHHKTIIVWLDLPIVQIIRRPK-HVVVVFVVFVFPDHTQLOEEL 131
DB 171 LNLLEKDYFGLTYRAENQKNWLDPAKEIKQVRSAGWH--FSFNVKFPYPPDPAQLSEDI 228

QY 132 TRYLFALQVQKDLAOGRLTCNDTSAALLISHIVQSIGDFD-EALDREHLAKNKYIPQO- 189
DB 229 TRYILCLQURDDIVSGRLPCSFVTLLALLGSYTVQSELGDDYDDEGSDYISFREFAPNHT 288

QY 190 DALEDKIVFEHHNHIGOTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGI 249
DB 289 KELEDKIVIELHSHRGMTPAEAMHFLFNAAKLSMYGVDLHAKDSEGVIMLGVCASGL 348

QY 250 LVFOGFTKINAFNNAKVRKLSFKRRLIKLRPDANSAYODTLEFLMASRDFCKSPWKIC 309
DB 349 LIYDRRLINRFAMPKVLKISYKRNIFYIKIRPGEPEQFESTIGFKLPNRAAKRLWKVC 408

QY 310 VEHAFPRLEFPKPKPVLFSGRSFSGRTQKVLDYVKEGHHKVKQPERKHSKTH 369
DB 409 VEHTHTFRLL-LPEAPPKKFL-TLGSKFYSRGTOAQTTRASALIDRPAPYFERSSSKRY 466

QY 370 ST-RSLASQPTELNSVLQSOOSTSLTGTGGAESPGGSCRGKPKVYSGEPGSHSP 428
DB 467 TMSRSL-----DGEVGTGQYATNG-----ISQTNLIITV 496

QY 429 APRRSAGNKQADGAASAPTEBEEVVKDRTQOKSKPPQPPSTGSLTSPHLSSELSVNSQ 488
DB 497 TPEKK-----AEERDEEDKRRKGEVTP-----ISAIRH 527

QY 489 GGVPANVTLSPLNLPDTPKQASPLISPLNDQACPRTDDE-----DGRKRPRPTDKAYFI 544
DB 528 EGKSPGLGDSCLPSPPSHCAP-TSPTELRRRCRNDCKLFGYEPSRAEHLFGEPAL-- 584

QY 545 AKEVSTTERTYKLDLEVITSWFQ-----YSKEDAMPEALKSLI-----FPNFEPL 591
DB 585 --DSGPGRPYLGDDQVAFSYRQAGKGTTLFSLSQLPESFSLDDGGLSFPNLS-- 640

QY 592 HKFHTNFKLE 601
DB 641 --ETNLLPQ 647

RESULT 13
QJMB3
ID QJMB3 PRELIMINARY; PRT; 1105 AA.
AC QJMB3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE TYPE II BRAIN 4.1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=BRAIN;
RX MEDLINE=20267856; PubMed=10806359;
RA Yamakawa H., Ohara O.;
RT "Comparison of mRNA and protein levels of four members of the protein
RT 4.1 family: The type II brain 4.1/4.1B/KIAA0987 is the most
RT predominant member of the protein 4.1 family in rat brain.";
RL Gene 248:137-145(2000).
DR EMBL; AB032827; BAA90774.1; -
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_41; 1.
```

```
DR PRINTS; PR00935; BAND41.
DR SMART; SM00295; B41; 1.
DR PROSITE; PS00660; BAND_41_1; 1.
DR PROSITE; PS00661; BAND_41_2; 1.
DR PROSITE; PS0057; BAND_41_3; 1.
SQ SEQUENCE 1105 AA; 122451 MW; 94FICA93A83693C6 CRC64;

Query Match      13.3%; Score 728; DB 11; Length 1105;
Best Local Similarity 32.6%; Pred. No. 9.2e-44;
Matches 198; Conservative 97; Mismatches 225; Indels 88; Gaps 20;

QY 18 ENSGISTLBERG-----KPPPTPSGKLVSIKIQMLDDTQEAEPVQAPGKVLDDAVCNH 73
DB 92 QSSSSKLSRSLKIVKRP-----KSMQCKVILLDGSYTCDEKRSQGVLFQVCKVCEH 146

QY 74 NLVEGDYFLEFPDHHKTIIVWLDLPIVQIIRRPK-HVVVVFVVFVFPDHTQLOEELT 132
DB 147 NLEKDYFGLTYRAENQKNWLDPAKEIKQVRSAGWH--FSFNVKFPYPPDPAQLSEDI 204

QY 133 RYLFALQVQKDLAOGRLTCNDTSAALLISHIVQSIGDFD-EALDREHLAKNKYIPQO-D 190
DB 205 RYILCLQURDDIVSGRLPCSFVTLLALLGSYTVQSELGDDYDDEGSDYISFREFAPNHTK 264

QY 191 ALEDKIVFEHHNHIGOTPAESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGIL 250
DB 265 ELEDKIVIELHSHRGMTPAEAMHFLFNAAKLSMYGVDLHAKDSEGVIMLGVCASGL 324

QY 251 VFOGFTKINAFNNAKVRKLSFKRRLIKLRPDANSAYODTLEFLMASRDFCKSPWKICV 310
DB 325 IYDRRLINRFAMPKVLKISYKRNIFYIKIRPGEPEQFESTIGFKLPNRAAKRLWKVCV 384

QY 311 EHHAFPRLEFPKPKPVLFSGRSFSGRTQKVLDYVKEGHHKVKQPERKHSKTHS 370
DB 385 EHTHTFRLL-LPEAPPKKFL-TLGSKFYSRGTOAQTTRASALIDRPAPYFERSSSKRYT 442

QY 371 I-RSLASQPTELNSVLQSOOSTSLTGTGGAESPGGSCRGKPKVYSGEPGSHSP 428
DB 443 MRSRLDGSVSNHENHETLYMKDSMAAEVGTGQYAT-----KVIS----- 481

QY 429 APRRSAGNKQADGAASAPTEE--EEVVKDRTQOKSKPPQPPSTGSLTSPHLSSELSVN 486
DB 482 -----QTNLIITVTPKAEHEEVEDRRKKAETP-----VAAL 518

QY 487 SOGGVAPANVTLSPLNLPDTPKQASPLISPLNDQACPRTDDEDEGRKRPRPTDKAYFI 546
DB 519 RHEGKSPGHGSDCLPSPPSHLDP-PSPTELRRRCRNDCKLFGYEPSRAEHLFGEPAL-- 572

QY 547 EYSTTERTYKLDLEVITSWFQ-----SVSKEDAMPEALKSLI-----FPNFEPLHK 593
DB 573 ESDAQCKAYLGDDQVAFSYRQAGKGTTLFSLSQLPESFSLDDGGLSFPNLS---- 628

QY 594 EHTNFKLE 601
DB 629 -ETNLLPQ 635

RESULT 14
O43491
ID O43491 PRELIMINARY; PRT; 1005 AA.
AC O43491;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROTEIN 4.1-G.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98260860; PubMed=9598318;
RA Parra M., Gascard P., Walensky L.D., Snyder S.H., Mohandas N.,
```

RA Conboy J.G.;  
RT "Cloning and characterization of 4.1G (EPB41L2), a new member of the  
RT skeletal protein 4.1 (EPB41) gene family.";  
RL Genomics 49:298-306(1998).  
DR EMBL: AF027299; AAC16923.1; -.  
DR InterPro: IPR000299; Band\_4.1.  
DR Pfam: PF00373; Band\_41; 1.  
DR PRINTS; PR00935; BAND41.  
DR PROSITE; PS00660; BAND\_41\_1; UNKNOWN\_1.  
DR PROSITE; PS00661; BAND\_41\_2; 1.  
DR PROSITE; PS00657; BAND\_41\_3; 1.  
DR SMART; SM00295; B41; 1.  
SQ SEQUENCE 1005 AA; 112587 MW; E86CB17488F6045F CRC64;

Query Match 13.2%; Score 720; DB 4; Length 1005;  
Best Local Similarity 31.2%; Pred. No. 3e-43;  
Matches 192; Conservative 102; Mismatches 222; Indels 100; Gaps 18;

QY 34 TPSSGLVSIKIQMLDDTQAFVPPORAPGKVLDDAVCNHLNVEGDFYFGLFEPDHKKITV 93  
Db 212 TKTKTVQCKVLLDGTETSCDEKHAQGVLFDKVCEHLNLEKDYFGLLFQESPEQKN 271  
QY 94 WDLPLPIVQIRRPKHVVVVFVFFPPDHTQLQBELTRYLFALQVQKODLAGRLTCDN 153  
Db 272 WLDPAKEIKRQLRNLPLWLT-FNVKFPDPSPQLTEDITRYFLCLQRODIASGRLPSCF 330  
QY 154 TSAALLSHIVQSEIGDFEALDREH---LAKNKVIP-QQDALEDKIVEFFHHNIGQTP 208  
Db 331 VTHALLGSTYLAQLGDYD---PEERHSIDLSEFQAPTQTKLEKVAELHKTHTGLSP 387  
QY 209 AESDFOLLEIARLEMYGIRLHPAKDREGTKINLAVANTGILVFOGFTKINAFNWAQVKR 268  
Db 388 AQADSQFLENAKRLSMYGVLDLHAKDSEGVQIKLVCANGLLIYKDRILNRNFAWPKILK 447  
QY 269 LSPKRRFLIKLRPDANSAYQDTLEFLMASRDFCKSFVKICVEHHAFFRFLFEPKPKPK 328  
Db 448 ISYKRSNFYIKVRPAELEQFESTIGFKLPNHRAAKRLWKVCVEHHTFYRLV-SPEQPPKA 506  
QY 329 VLFSSGSSFRFSRTQKQVLDYVKEGGHKVQPERKHSKIHRSIRLASQTELNSEVLEQ 388  
Db 507 KFTLGSKFRYSRTQAQTRQATSLDIRPAPHFRTSSKRVIS-RSLDGAPI---GVMDQ 561  
QY 389 SQOSTSLTTEGAEPSGGSCRRGKPKVSAGPSPAPRRSPAGNKQADGAASAP- 447  
Db 562 SLMK-----DFPGAAG-----EISAYGPGI-VSIAYVQDGDGRREVRSPTKAPH 604  
QY 448 -----TEEEEVVKDRT-----T-----QOSKPPPP- 467  
Db 605 LQIEGKNSLRVEGDNIVYRHSNLMLELDKAQEDILKHOASISELKRNFMESTPEPRP 664  
QY 468 -----OPSTGSLTSPH---LSLSVNSOGGVAPANVT-----LSPNLS-PTK 507  
Db 665 NEWEKRITPLSLQTCGSHETLNIVEEKKRAEVGKDERVITEEMNGKELSPGSGGEIR 724  
QY 508 QASPLI-----SPLNDQACPRTDDEGRKRKRPPTKAYFIAKEVSTTERTYTKLQLEVIT 563  
Db 725 KVEPVTKDSTLSLSSSSSSSESEEDVGEYRPHR---VIEGTIRREQVEVEVEEPP 781  
QY 564 SWFQSTVSKEDAMPEA 579  
Db 782 RPAKVVEREEAVPEA 797

RESULT 15  
Q9R102 PRELIMINARY; PRT: 706 AA.  
ID Q9R102;  
AC Q9R102;  
DT 01-MAY-2000 (TREMBlrel. 13, Created)  
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
GN DALI.

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C; TISSUE=BRAIN;  
RA Azam M., Andrabli S., Lin L., Newsham I., Chishti A.H.;  
RT "Mouse DAL-1 (MDAL-1) cDNA Sequence."  
RT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF177146; AAD51365.1; -.  
DR InterPro; IPR000299; Band\_4.1.  
DR Pfam; PF00373; Band\_41; 1.  
DR PRINTS; PR00935; BAND41.  
DR PROSITE; PS00660; BAND\_41\_1; 1.  
DR PROSITE; PS00661; BAND\_41\_2; UNKNOWN\_1.  
DR PROSITE; PS00657; BAND\_41\_3; 1.  
DR SMART; SM00295; B41; 1.  
SQ SEQUENCE 706 AA; 79448 MW; 75B6A3EF475A5E31 CRC64;

Query Match 13.0%; Score 711.5; DB 11; Length 706;  
Best Local Similarity 32.4%; Pred. No. 7.5e-43;  
Matches 185; Conservative 99; Mismatches 250; Indels 37; Gaps 12;

QY 43 KIQMLDDTQAFVPPORAPGKVLDDAVCNHLNVEGDFYFGLFEPDHKKITVWLDLKP 102  
Db 4 KVTLLDSEYGVCDVRSRGQVLFDKVCEHLNLEKDYFGLTYRDAENKKNWLDPAKEIK 63  
QY 103 KOIRRPK-HVVYKVVVVFVFFPPDHTQLQBELTRYLFALQVQKODLAGRLTCDN 161  
Db 64 KOIRSGAWH-FSFNVKFPDPAPQLSEDTIRYVLCQLRDDLVSGLRPFVTLALGS 121  
QY 162 HIVQSIGDFD-EALDREHLAKNKYIPQ-QDALEDKIVEFFHHNIGQTPAESDFOLLE 219  
Db 122 YTVQSELDYDDECCNDYISEFRFAPNHTKELEDKVIELHKSHRGNTVPEAEHMFLENA 181  
QY 220 RLEMYGIRLHPAKDREGTKINLAVANTGILVFOGFTKINAFNWAQVKRSLPKRRFLIK 279  
Db 182 KKLMSYGGDLHAKDSEGVIMLGVCASGLLIYRDLRINRFAWPKVLSYKRNIFYK 241  
QY 280 LRPDANSAYQDTLEFLMASRDFCKSFVKICVEHHAFFRFLFEPKPKPKVLFSSGSRFR 339  
Db 242 IRPGEQFESTIGFKLPNHRAAKRLWKVCVEHHTFRLI-LPEAPPKFL-TLGSKERY 299  
QY 340 SGRTOQVLDYVKEGGHKVQPERKHSKIHSI-RSLASQ---PTELNSEVLEQSQOSTSL 395  
Db 300 SGRTOQVTRASALIDRPAPYFERSSSKRYTMSRSLDGEVGTGQYATTGISQTNLITV 359  
QY 396 TFGEGAEPSGGSCRRGKPKVSAGPSPAPRRSPAGNKQADGAASAPTEEEVEV 455  
Db 360 TPEKKAEEB-----RVEEEDRRKKAEEATPVIALRHEGKTDSERTDAADGETSATEDLD 414  
QY 456 KRTQOSKPPQPPQSTGSLTSPHLSLSVNSOGGVAPAN-----VTLSPNLSPTDKQAS 510  
Db 415 KQDELMLKHQ-----TNISELKRTELETSTALTNEWKRLSTSPVRLAARQEDA 465  
QY 511 PLISPLNDQACPRTDDEGRKRKRPPTKAYFIAKEVSTTERTYTKLQLEVITSWFQSTV 570  
Db 466 PMIEPLVPEEKLETKTEPVEAEVESTPHPOPL-----STEKVLOETILVEERHVMVSH 518  
QY 571 SKEDAMPEALKSLIFPNFEPHLKHFHTNFKLE 601  
Db 519 ASGDASHARTARDEVDAAESTPTDRRHTGKGKE 549

Search completed: December 6, 2001, 08:54:30  
Job time: 272 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2001, 12:06:44 ; Search time 4659.49 Seconds  
(without alignments)  
11099.639 Million cell updates/sec

Title: US-09-555-342A-1\_COPY\_49\_3183

Perfect score: 3135

Sequence: 1 atggggaataagacagag.....acaaagatctctgtgtat 3135

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_htg.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vl.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_om.\*

20: em\_or.\*

21: em\_ov.\*

22: em\_pat.\*

23: em\_ph.\*

24: em\_pl.\*

25: em\_ro.\*

26: em\_sts.\*

27: em\_sy.\*

28: em\_un.\*

29: em\_vl.\*

30: em\_htgo\_hum.\*

31: em\_htgo\_inv.\*

32: em\_htgo\_rod.\*

33: em\_htg\_hum.\*

34: em\_htg\_inv.\*

35: em\_htg\_rod.\*

36: em\_htg\_other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	3135	100.0	3442	9	AB008430	AB008430 Homo sapi
2	1043	33.3	2632	10	BC004009	BC004009 Mus muscu
3	1037.4	33.1	3997	9	AB018336	AB018336 Homo sapi
4	1013.2	32.3	3719	10	BC009153	BC009153 Mus muscu
5	342.8	10.9	2888	9	HSN802608	AL161984 Homo sapi
6	286.6	9.1	1438	9	HSN801276	AL122052 Homo sapi
7	254	8.1	96183	9	AL161896	AL161896 Human DNA
c	254	8.1	152053	2	AC022669	AC022669 Homo sapi
8	254	8.1	181800	2	AL137249	AL137249 Homo sapi
9	210	6.7	483	11	G22662	G22662 human STS W
c	194	6.2	2595	9	AF156225	AF156225 Homo sapi
11	193.4	6.2	152053	2	AC022669	AC022669 Homo sapi
12	193.4	6.1	2867	9	HUMELIA	M14993 Human struc
13	192.4	6.1	3043	9	HUMEMP41	M61733 Homo sapien
14	192.4	6.0	5942	10	MUS41SP	L00919 Mus muscullu
15	189.4	6.0	2882	10	AF106702	AF106702 Mus muscu
16	187	5.9	6102	10	AB019256	AB019256 Rattus no
17	185.2	5.9	5855	10	AB019257	AB019257 Rattus no
18	183.6	5.9	3619	9	BC010674	BC010674 Homo sapi
19	181.8	5.8	3643	9	HUMPTYPH	M68941 Human prote
20	181.2	5.8	2590	4	AF222767	AF222767 Bos tauru
21	181.2	5.8	3661	10	AB032828	AB032828 Rattus no
22	181.2	5.8	4543	10	AB032827	AB032827 Rattus no
23	181.2	5.8	6263	9	AB002336	AB002336 Human mRN
24	180.4	5.7	5921	3	DROP41A	L27467 Drosophila
25	179.6	5.6	2964	10	AF044312	AF044312 Mus muscu
26	176.6	5.6	3320	10	AF061283	AF061283 Mus muscu
27	174	5.5	170916	9	AL136300	AL136300 Human DNA
28	172.8	5.5	2758	5	XELCSK	M20621 X.laevis cy
29	172.4	5.4	4336	9	AF027299	AF027299 Homo sapi
30	169.6	5.4	3451	10	AB032366	AB032366 Mus muscu
31	166.8	5.3	4051	10	AF152247	AF152247 Mus muscu
c	165	5.3	393	11	G60059	G60059 SHGC-130885
32	162.4	5.2	2717	10	AF177146	AF177146 Mus muscu
33	161.8	5.2	3309	9	AF069072	AF069072 Homo sapi
34	161.6	5.2	3387	9	BC006141	BC006141 Homo sapi
35	161.6	5.2	4446	9	AB023204	AB023204 Homo sapi
36	161.6	5.2	1973	9	AB070172	AB070172 Macaca fa
37	157.6	5.0	2677	9	HUMELI	J03796 Human eryth
38	154.2	4.9	42014	2	AC018327	AC018327 Drosophil
39	154	4.9	160710	3	AC009537	AC009537 Drosophil
40	154	4.9	164713	3	AC007589	AC007589 Drosophil
41	154	4.9	303209	3	AE003604	AE003604 Drosophil
42	154	4.6	3984	9	HUMCAP	M64572 Human prote
43	143.2	4.6	2211	10	BC011476	BC011476 Mus muscu
44	143.2	4.6	2211	10	BC011476	BC011476 Mus muscu
45	143.2	4.6	2211	10	BC011476	BC011476 Mus muscu

#### ALIGNMENTS

#### RESULT 1

AB008430

LOCUS

DEFINITION

AB008430

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

1 (bases 1 to 3442)

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (22-OCT-1997)

to the DDBJ/EMBL/GenBank databases. Takeshi

Kawamoto, Hiroshima University School of Dentistry, Department of

Biochemistry; 1-2-3 Kasumi Minami-ku, Hiroshima, Hiroshima 734,

Japan (E-mail:tkawamo@ipc.hiroshima-u.ac.jp, Tel:082-257-5688,

Fax:082-257-5629)

2 (sites)

AUTHORS Koyano, Y., Kawamoto, T., Shen, M., Yan, W., Noshiro, M., Fujii, K. and Kato, Y.

TITLE Molecular cloning and characterization of CDEP, a novel human protein containing the ezrin-like domain of the band 4.1 superfamily and the Dbl homology domain of rho guanine nucleotide exchange factors

JOURNAL Biochem. Biophys. Res. Commun. 241 (2), 369-375 (1997)

MEDLINE 98086358

FEATURES

Source

Location/Qualifiers

1..3442

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/cell\_type="chondrocyte"

/dev\_stage="embryo"

/tissue\_type="cartilage"

49..3186

/function="rho Guanine Nucleotide Exchange Factor"

/note="Band 4.1 superfamily"

/codon\_start=1

/product="CDEP"

/protein\_id="BAA24267.1"

/db\_xref="GI:2766165"

/translation="MGEIEQRTPGSRIGCAPENSGISTLERQKPPPTSGKLVSIKI QMDDTFAFVQAPQKVLVDVQCNHLNVLGYDFGLEFPDHRKKTIVWLDLLKPIV KQIRPKHVVVYKFFPDHDTQLQELRYLFALQKDLAQGLRGTCTNDTSAALLI SHVQSEIGDFDEALDRHLAKNKV IPOODALEDKIVEFHNIHIGOTPAESDFOLLEI ARLEDMYGRHLPAKDRECTKINLAVANTGILVFOGFTKINAFNAKVKFLSKRFR LKLRPDANSAYODTLEFLMASRDFCKSPWKICVSEHAFPRLEPEPKPKPVLSRG SFRSGRTQKQVLDIVKGGHKKVQFERKSHKIHISIRLSQPTLELSEVLEQSQS TSLTGEAGSEGGSCRRGKPKVSAGEPGSHPSAPRRSPAGNKQADGAAPTEE ESEVVKDFTQKQPPQPSGTSLGSPHLSVNSOGGVPANVTLSPLSPDTKE AEVLPSLINDQACPTDDEDEGRKRFPTDKAYTIAKEVSTERTYLLKDLVITSWF QSTVSKEDAMPALKSLIFPNPEPLKHTNFKLEORLALWEGRSNAQIRDYORIG DVMKNIQGMKHLAHLKHSALALENGIKSRRLNFCRDFELQKVCYLPILNFTL LRLRLMHYQVLERLCKHPPSHADFDRCRAALAEITEMVAQHLGTMKNENFOLK HEKQDLIGIDNVVPGREFIRGLSKSLGKLOQRMFFLFNDVLLTSLRGLTASNQ FVKHQGLPYGMTIESEDEWGPVHCLTLRGQRSIIIVAASRSRSEMKWVEDIQMAID LAEKSSPAPELASSPPDNKSPDEATAADQESDLDASRTSLERQAPHRGNTWVHV CWHRTSVSMVDFSLAVENQLSKNLLRKFKNNGWKLWVFTNFCLEPYKSHQDNHP LASPLPLGLSLTIPSESENIQKDYVFKLHFKSHVYFFRAESYTERWMEVIRSATSS ASRPVLSHKSLESLV"

polya\_site

3442

/note="50 a nucleotides"

BASE COUNT 864 a 952 c 927 g 699 t

ORIGIN

Query Match 100.0%; Score 3135; DB 9; Length 3442;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 3135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgggagaaatagacagagcgcccccagatcacgactgagggcgcccggaataatcg 60

Db 49 ATGGGAGAAATAGACAGAGCGACCCAGGATCACGACTGGGGGCCCCGGAAATTCG 108

QY 61 gggatcagctaccttggaacgtggacagagcgcccccacaccttcaggaaactcgtg 120

Db 109 GGGATCAGTACCTTGGAACTGGACAGAGCGCCGCCACACCTTCAGGAAACTCGTG 168

QY 121 tccatcaaatccagatgctggtgatgacacccagagcgatttgaagttccacaaagact 180

Db 169 TCCATCAAAATCCAGATGCTGGATGACACCCAGAGGCGCATTTTGAAGTTCACAAAGAGCT 228

QY 181 cctgggaagtgcctgctgagtcagttgcaaacacctcaacctcgtgaagtgactat 240

Db 229 CTTGGGAAGTGCCTGCTGATGTCAGTTTGCACACCTCAACCTCGTGGAAAGTGACTAT 288

QY 241 ttggcctcagtttctctgatacaaaaagatcacggtggtggtgagatctcctaaaccc 300

Db 289 TTTGGCCTCGAGTTTCTTGATCACAATAAAGATCACGCTGTGCTGATCTCTTAAACCC 348

QY 301 attgtgaacagattagaaggccaaagcagctgtgtgttaagtgtgtgtaattcttt 360

Db 349 ATTGTGAAACAGATTAGAAGCCCAAGACAGTGTGTTGTTAAGTTTGTGGTGAATTCCTTT 408

QY 361 cgccttgaccacacacacaaactccaaagaactcacaaggttaacctgttcgcgtgcaggtg 420

Db 409 CCCCTTGACCACACAACTCCCAAGAAGAACTCACAAAGGTACTGTTCGCCCTGCAAGTG 468

QY 421 aagcaggaacttggtcctcaagcgaggttgacgtgtaataagacaccagcgagctctcttgatt 480

Db 469 AAGCAGGACTTGGCTCAAGCGCAGGTTGACGTGTAATGACACACGCGCAGCTCTCTTGATT 528

QY 481 tccacattgtgcaatctcagatttgggatttggatgaaagccttgacagagagcactta 540

Db 529 TCACACATTGTGCAATCTCAGATTGGGGATTTTGATGAAGCTTTGGACAGAGACACTTA 588

QY 541 gcaaaaaataataacataactcagcaagacgacactagagagcaaaaatcgtgggaattcac 600

Db 589 GCAAAAAATAATATACATACCTCAGCAAGACGCACTAGAGGACAAATTCGTGGAAATTCAC 648

QY 601 cataaccatttggaacacacaccagcagaatcagatttccagctccttagagattgcccgt 660

Db 649 CATAACCACATTGGCAACACACCCAGCAGAATCAGATTTCACGCTCTAGAGATTGCCCGT 708

QY 661 cggctagagatgtatggaatccoggttgccccggcccaagagacaggaagacagaagatc 720

Db 709 CGGCTAGAGATGTATGGAATCCGGTTGCACCCGCCCAAGACAGGGAAGCACGAAGATC 768

QY 721 aactcgccgttgccaaacgggaattctagttttcaggggtttcactaagatcaatgcc 780

Db 769 AATCTGGCCGTTGCCAACACGGGAATTTCTAGTGTTCAGGGTTTCAATAGATCAATGCC 828

QY 781 ttcaactgggccaaggtgcggaagctgagcttcaagaggaagcgcttctcatcaagctc 840

Db 829 TTCAACTGGGCCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGCTC 888

QY 841 cggccagatgccaatagtgcgtaccaggtatccttggaattcctcgtatgcccagtcgggat 900

Db 889 CGGCCAGATGCCAATAGTGTGCTACAGGATACCTTGGAAATTCCTGTATGGCCAGTCGGGAT 948

QY 901 tctcgaagctctcttgaaaatctggttgaaacatcagctcctcttttagactttttaa 960

Db 949 TCTGCAAGTCTCTTGGAAAATCTGTGTGAACATCATGCTCTTTAGACTTTTGTAA 1008

QY 961 gagcccaaaccaagccccagccgctcctcttttagccgggggttcacatttccggttcagt 1020

Db 1009 GAGCCCAACCAAGAACCCAGCCCGTCTCTTTAGCCGGGGTCAATTTTCGTTCACT 1068

QY 1021 ggtcgactcagaagcaggttctcgaactagttaaagaagagagacataagaaggtgcag 1080

Db 1069 GGTGGACTCAGAAGCAGGTTCTCGACTATGTATAAAGAGGAGGACATAAGAAGGTGCAG 1128

QY 1081 ttgaaagaagcacacagcagattcattccggagccttgcttcacagcctacagaa 1140

Db 1129 TTTGAAAGAGACACAGCAAGATTCTATCTATCCGGAGGCTTGTCTTCACAGCCTACAGAA 1188

QY 1141 ctgaattcgggaagtgcgtgagcagctcagcagagcaccagccttaccatttggagaaggt 1200

Db 1189 CTGAATTCGSAAGTGTGGAGAGTCTCAGCAGAGTCTCAGCAGACCCAGCCCTTACATTTGGAGAAGGT 1248

QY 1201 gccgaatctccagggggccagagctgcgcgaggaagaaagacccgaaggtttccgcgag 1260

Db 1249 GCCGAATCTCCAGGGGGCCAGAGTCTCCGGCGGAGGAAGAACCCGAAGGTTTTCGCCGGGG 1308

QY 1261 gagcggggtgcacacccagcctgcgcgagagagagccccgcggttaacagcagcg 1320

Db 1309 GAGCCGGGTTCGACCCGAGCCCTTCGCCGAGGAGAGGCCCGCGGTAAACAGCAGCGC 1368

QY 1321 gagcgagcgctcgcgcgcccaagagagagagagaggttcgttaagatagagacccag 1380

Db 1369 GACGGAGCCGCTCGCGCCCCACGGAGGAAGAGGAGGAGGTCTGTTAAGGATAGGACCCAG 1428

QY 1381 cagagttaacctcagccccccgcagccaagcacaaggtcctcactgagcagctcctacctt 1440

Db 1429 CAGAGTAAACCTCAGCCCCCGCAGCCCAAGCACAGGCTCCCTTGACTGGCAGTCTCCTACCTT 1488

QY 1441 tccgagctgtctgtgaactcgccaggggggagtggtgcccctgcccacacgtgaccttgytctccc 1500  
Db 1489 TCCGAGCTGTCTGTGAACTCGCGAGGGGAGTGGCCCTGCCAACGTGACCTTGTCTCCC 1548  
QY 1501 aaactgagccccacaccacagagcctctcccttgaatcagcccgctgtgaaatgaccag 1560  
Db 1549 AACCTGAGCCCCGACACCAACAGCAGGCTCTCCCTTATCAGCCCGCTGCTGAAATGACCAG 1608  
QY 1561 gccctgccccggagcgacgatgaggatgagggccggaggaagagattcccaactgataaa 1620  
Db 1609 GCCTGCCCCGGACGACGATGAGGATGAGGCGCGGAGGAAGAGATCCCAACTGATAAA 1668  
QY 1621 gctacttaactaagaaagtgtctaccacgagcgaacatatctgaaagatctctcgaa 1680  
Db 1669 GCCTACTTTCATAGCTAAGGAAGTGTCTACCACCGAGCGAACAATATCTGAAGGATCTCGAA 1728  
QY 1681 gttatcacttcgtggtttcagagcacagtgagcaagaagagacccatgcccgaagcactg 1740  
Db 1729 GTTATCATTCTGTTGTTTCAGACACAGTGAAGCAAGAGGACCCCATGCGCGAAGCACTG 1788  
QY 1741 aaagtctcatattcccgaaattttgaacctttgcacaaatttcatataatttttccaag 1800  
Db 1789 AAAAGTCTCATATTCCCGAATTTTGAACCTTTGACAAATTTTCATACTAATTTTCTCAAG 1848  
QY 1801 gaaattgagcaacgaacttgccctgtgtgggaaggccgtccaaatgcccaaatcagagattac 1860  
Db 1849 GAAATTGAGCAACGACTTGGCCCTGTGGGAAGGCGCGCTCAAAATGCCAAATTCAGAGATTAC 1908  
QY 1861 caaagaatcgcgatctatgctgaagaacattcagggaatgaagacacctggcggtctcac 1920  
Db 1909 CAAAGAATCGGCCGATGTCACTGTGAAGAACATTCAGGGCATGAAGCACCTTGGCGGGCTCAC 1968  
QY 1921 ctgtggaagcacagcagggccttggaggccctggagaatggaatcaagagctcccgcgcgg 1980  
Db 1969 CTGTGAAGCACAGCGAGGCGCTTGGAGGCGCTGAGAGATGGAATCGAATCGAAGCTCCCGGCGG 2028  
QY 1981 ctggagaactctgcagagactttgagctgcagaaggtgtgttacetaccgctcaacacc 2040  
Db 2029 CTGGAGAACCTTCTGCAGAGACTTTGAGCTGCAGAAGGTGTGTACCTACCGCTCAACACC 2088  
QY 2041 ttccctctcgccactgcaccggtcatgtcactacaagcagctcctgagcgcgtgtgc 2100  
Db 2089 TTCTCTCTGGGGCACTGCAACCGGCTCATCGCTCAAGCAGGCTCTGGAGCGGCTGTGC 2148  
QY 2101 aaacacacccgcgagccagccgacttcaggagactgcgagccgctttggcagagatc 2160  
Db 2149 AAACACACCCCGGAGCCGACGCGGACTTCAGGACTGCGGAGCGCTTTGGCAGAGATC 2208  
QY 2161 acggagatggtgacagctccacgtaacgatatacaagatgagaaatttccagaagctg 2220  
Db 2209 ACGGAGATGTTGTCACAGCTTCCACGGTACCGATGATCAAGATGAGAAATTTCCAGAAAGCTG 2268  
QY 2221 cagaaactcaagaatttgatggcattgacattgacaattctgtgttcgggaaggagttc 2280  
Db 2269 CAGCAACTCAAGAAGATTGTATGGCATTTGACAAATCTTGTGTTCCGGGAAGGGAGTTC 2328  
QY 2281 atccgtctggcgacctcagcaagctctcgggaagggtcccgagcagcagctgttcttc 2340  
Db 2329 ATCCGCTGGGCAGCTCAGCAAGCTCTCGGGGAAGGGCTCCAGCAGCGCATGTTCTTC 2388  
QY 2341 ctgttcaacgaactctctgtatatcacagcgcgggggtgagcgccctccaatcagtttaa 2400  
Db 2389 CTGTTCAACGACGCTCTGCTATACAGAGCGGGGGCTGACGGCTCCCAATCAGATTAAA 2448  
QY 2401 gtccacgggcagctcccgctctatgcatgacatgagagagcgaagacgagtggggg 2460  
Db 2449 GTCCACGGGCAGCTCCCGCTCTATGGCATGACGATTGAGGAGAGCGAAGCAGTAGTGGGG 2508  
QY 2461 gtgccccactgctgacctccggggcagcgcgcagtcctcatcgtggccgcagctct 2520  
Db 2509 GTGCCCCACTGCTGACCTTCCGGGGCCAGCGGCGAGTCCATCATCTGTCGGCGCCAGTTCT 2568

QY 2521 cgtccgagatgagaagtgggttggaggacatccagatggccattgaccttgccggagaag 2580  
Db 2569 CGGTCCGAGATGAGAAGTGGGTGTGAGACATCCAGATGCCATTGACCTGGCGGAGAAG 2628  
QY 2581 agcagcagccccccctcgtgattccttgccagcagcagccccctgacaacaagtccccctgat 2640  
Db 2629 AGCAGAGCCCCCCTGAGTTCTTGGCCAGCAGCCCCCTGACACAGTAGTCCCTTGAT 2688  
QY 2641 gaagccacgcggcgtgaccaggagtcagagagtgacctgagcgcctcgccacatcgctg 2700  
Db 2689 GAAGCCACCGCGGTGACACGAGGTGAGAGATGACCTGAGCGCTCGCCACATCGCTG 2748  
QY 2701 gacgcccaggccccgcaccgcggaacacaaatggtgcaactgtgctgacccgcaacacc 2760  
Db 2749 GAGCGCAGCGCCCGCACCGCGCAACAATGTTGACCTGCTGGCACCTGCTGCGCAACCC 2808  
QY 2761 agcgtctccatggtgacttcagcatcgagtcagagaaatcagttgtctggaacactgctg 2820  
Db 2809 AGGTCTCCATGTTGACTTCAGCATCGCAGTGGAGAAATCAGTTGTCTGGAACCTGCTG 2868  
QY 2821 aggaatttcaaaaacagcaagcgggtggcagaagctgtggtgtgttcacaaaacttctgc 2880  
Db 2869 AGGAAATTTCAAAAACAGCAACGGGTGGCAGAAGCTGTGGTGTGTTCACAAAATCTCTGC 2928  
QY 2881 ctgttcttctacaaatcacaccaggacaatcatcccttccagcctgctctgctgcgc 2940  
Db 2929 CTGTTCTTTACAAATCACACCAGGACAATCATCCCTTCCAGCCTGCTCTGCTCGGC 2988  
QY 2941 tactcgtccacctccctctgagtcgagaaatccagaaagactacgtgttcaagctg 3000  
Db 2989 TACTCGCTCAACATCCCTCTGAGTCCGAGAACAATCCAGAAAGACTACGTGTTTCAAGCTG 3048  
QY 3001 cacttcaagtccacactctactacttcaaggcggaaagcaggtacacgttctcgaaggtg 3060  
Db 3049 CACTTCAAGTCCCACTACTACTTCAAGGCGGAAAGCGAGTACAGCTTCGAAAGGTGG 3108  
QY 3061 atggaagtatccgcagtcgccaccagctctgctcgaccccaactggtgagccacaaa 3120  
Db 3109 ATGGAAGTATCCGCAGTGCACACAGCTCTGCTCGAGCCCCACACGTTGTGAGCCACAAA 3168  
QY 3121 gagtctcttctgtat 3135  
Db 3169 GAGTCTCTGTGTAT 3183

RESULT 2  
BC004009  
LOCUS BC004009 2632 bp mRNA ROD 12-JUL-2001  
DEFINITION Mus musculus, clone IMAGE:3493093, mRNA, partial cds.  
ACCESSION BC004009  
VERSION BC004009.1 GI:13278387  
KEYWORDS  
SOURCE house mouse.

ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE  
1 (bases 1 to 2632)  
AUTHORS Strausberg, R.  
TITLE Direct Submission

JOURNAL  
Submitted (28-FEB-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk

COMMENT  
Email: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Center, Stanford University School of Medicine, Stanford, CA 94305  
Web site: <http://www-shgc.stanford.edu>  
Contact: (Dickson, Mark) [mcd@paxil.stanford.edu](mailto:mcd@paxil.stanford.edu)



REFERENCE 2 (bases 1 to 3997)  
 AUTHORS Ohara,O., Suyama,M., Nagase,T., Ishikawa,K. and Kikuno,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-OCT-1998) to the DDBJ/EMBL/GenBank databases. Osamu Ohara, Kazusa DNA Research Institute, Laboratory of DNA Technology; Yana 1532-3, Kisarazu, Chiba 292-0812, Japan  
 (E-mail:cdnainfo@kazusa.or.jp, Tel:+81-438-52-3913, Fax:+81-438-52-3914)

## FEATURES

source location/Qualifiers  
 1. 3997

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="hk05692"

/clone\_lib="pBluescriptII SK plus"

/dev\_stage="adult"

/sex="male"

/tissue\_type="brain"

gene 118..3282

/gene="KIAA0793"

CDS 118..3282

/gene="KIAA0793"

/codon\_start=1

/product="KIAA0793 protein"

/protein\_id="BAA34513.1"

/db\_xref="GI:3882307"

/translation="MGEIEGTVRVLTQAGRLGAQTPVGVSTLEPGQTLPRMQEKHL  
 HLRKLDLNTWEIFDIEPKDQVLLTQWRLNLVECDYFGMEFQNTQSYWIWLEPM  
 KPIIRRRPNVLRVLAFFPPDPGQLEQYRYLFALQKRLDLRLTCAADTTA  
 ALLSHQLQSIGDYDETLDRHLKVNELPQOHLKLEKILEFHOKHVGOTPAESDFQ  
 VLEARKLEMGVIRPHMASDREGKIQIOLAVSHMGVLVFGQTKINTFNWSVKRLSFK  
 RKRFLIKLPEVHPQYDTLEFLGSRDECKFNWKCIVEYHTFRLQLQPKKARVF  
 FRSRSSYRGKQLVDYFKDSMKRPIYERRSKHTHTSVRLATDLPKOSISFFE  
 GLRTPASSANZUYSLSPLVPGLPEFKDSSSLDPOVSYVKSPPAAERRSGAVA  
 GQPTPSAGLPALQPGPLSKSPQSPSRKSLSPAFQVPLGPAEQSSPL  
 LSPVLSDAGGAGCEEPERHVRPADEAYFVKEILATERYLDLEVITVWFRSAV  
 KEDAMPATLLLESNIDPIYEFHRGFLREVEORLALWEGPSKATGSHORIGDILL  
 RNMRQLEFTSYFQRHDSVLELEKATRCCKKLEAYKEFELOKVCYLPILFTLLKPI  
 QRLHYRLLLRGLCHYSPGHVDYADCHDLKAITEVTTLOHILIRLNLQKULEPQ  
 RDLVGIENLAPRGFTREGCLHLKLTQKQRMFFLFSDMLLYTSKVGATSHFRIR  
 GILLPLQGLMVEDSNESWPHCFITYAAQKTVVAASRLEKEKWMKLDLSAIOAKS  
 GGDVAPALPTVCTRPSPRSPNEVLEQESDDARGVRSLEGHGHRANTTMHVCWY  
 RNTSVSRADHSAANVENOLSGYLLRKFNKSHGOKLWVFTNFCFLFYKTHODDYPLAS  
 LPLIGYSVSIPEADGIHKHVKFLQFKSHVFFRAESKYTFERWMEVIQGASSAGR  
 APSIVQDGPQSPSLGLEGVMVRGKEE"

BASE COUNT 998 a 1132 c 1068 g 799 t

## Query Match

33.1%; Score 1037.4; DB 9; Length 3997;

Best Local Similarity 60.2%; Pred. No. 2.9e-221;

Matches 1864; Conservative 0; Mismatches 1186; Indels 45; Gaps 7;

Qy 21 gccacccagatcacagctggggccggccgaaattcgggagatcagcttggaagc 80

Db 150 GCAGACTGCGGGATGCGCTTGGGTGCCAGACCCCTGTGGGAGTTAGCACCCCTGAGCC 209

Qy 81 tggacagaagcgcgcccaacaccttcagaaaaactcgtgtccatacaaatccagatgct 140

Db 210 TGGCAGACTCTCTGCCAGAAATCAAGAGAGACACCTGCACCTCAGAGTAAAGCTGCT 269

Qy 141 gaatgacacccagggagcatttgagttccacaaagagctccttggaagtgctgctgga 200

Db 270 GGACACACCATGGAATATTTGACATTGAGCCTAAATGCGATGCGCAGGTATTACTGAC 329

Qy 201 tgcagttgacacacacctcaactcgtggaagtgactattttggcctcgagtttctga 260

Db 330 ACAAGTGTGGAGCGTTTAAACCTGGTAGAATGTGACTACTTCGGGATGGAGTTTCAAAA 389

Qy 261 tcacaaaaagatcaacggttggtggtgattcctcaaaacccattgtgaaacagattagaag 320

Db 390 TACTCAGTCTTACTGGATTGGCTTGAACCTATGAACCCCATCATTTAGCAAAATACGAG 449

Qy 321 gccaaagcagctgtgttaagtgttggtgaaattctttccgcctgaccacacaaact 380

Db 450 GCCAAGAATGTGGTCTTCGGCTAGCTGTAAAATTTTTTCCACCTGATCTCTGGTCACT 509

Qy 381 ccaagaagaactcacaaagttacctgttcgcgtgcaggtggaagcagcagcttggtctcaag 440

Db 510 ACAAGAAGAATATACAAAGTACTTGTTCGCTTGCAACTTAAGAGAGACCTGCTGGAAGA 569

Qy 441 csaggtgacgtgtaataagcaccagcagcagctctctgtattcacacatgtgcaaatctga 500

Db 570 GGGTTTGACCTGTGCTGACACACACAGCGCCCTCTTCACGTCCCTCTCTCGAGTCGA 629

Qy 501 gattgggagatttgatgaagccttgacagagagcacttagcaaaaaataataacataacc 560

Db 630 AATAGAGATTACGATGAACCGCTGGACCGAGAGACCTCAAAAGTGAACGAGTATTTGCC 589

Qy 561 tcagcaagacgcactagagacaaaaatcgtggaatttcaccataaccacacttggaacaac 620

Db 690 TGGCCAGCAGCAGCTGCTTGTGAGAAGATACTAGAAATTCATCAGAACGCTGGGCCAGAC 749

Qy 621 accagcagaatcagatttccagctcctagagattgcccgtcgcgctagagatgtatgaat 680

Db 750 ACCTGCTGAGTCGGATTTCAGGTGCTCGAAATTTGTCGAAAAGTTGGAATGTACGGCAT 809

Qy 681 ccggttgaccggcccaagacagaggaagcagagatcaatctgcccgttgcccaaac 740

Db 810 CAGATTTCACATGGCTTCTGACAGGAAGAACCAAGATTCAAACTGGCAGTTTCCCACAT 869

Qy 741 gggaaattcagtttccaggggttccactaagatcaatgccttcaactgggccaagtgcg 800

Db 870 GGGTACTCGTGTCCAGGGCACCACCAAAATCAACACTTTCACTGGTCCAAAGTCCG 929

Qy 801 gaagctgagcttcaagaggaagcgtcttctcatcaagctccggccagatgccaatagtc 860

Db 930 TAAACTAAGCTTCAAGAGGAAAAGATTCTTATCAAACTTCATCCAGAGGTTTCATGGACC 989

Qy 861 gtaccagagatacttggaattcctgaatgccagtcgggattctcgaactcctctggaa 920

Db 990 TTACCAGGACACATTAGAATTTTGTGGGTAGTAGAGATGAATTAAGAACTCTCGAA 1049

Qy 921 aatctgtgtgtaacacatcgtcttcttagactttttgaagcggcccaaaccaagccaa 980

Db 1050 GATTGTGTGGAGTATCACACCTTTTGTAGACTTTTGGACCAACCTTAAGCCAAACAAA 1109

Qy 981 gccgctcctcttagcgggggttcattcattcgtggttcagtggtcgagctcagaagcaggt 1040

Db 1110 AGCGGCTCTTCTTCAGCGGGGCTCCCTTCAGATACAGTGGGAAGAACTCAGAAACAACT 1169

Qy 1041 tctcgaactgttaagaagagagagacataagaaggtgcagtttgaaaggaagcagacaa 1100

Db 1170 AGTAGATTATTTCAAGACAGCTGGAATGAAGAGAAATTCATATGAAGAAGGCACAGCAA 1229

Qy 1101 gattcattctatccggagccttgcttcacagcctcacagaactgaattcgaagtgctgga 1160

Db 1230 GACCCACAGCTC-----GTTCCAGCTCTGACTGACGACCTTACCATA 1271

Qy 1161 gcagttcagcagagcaccagccttacatttgagagaggtgcagaaatctccagggggcca 1220

Db 1272 ACAGAGACTCTCATTCGCCGAGGGATTGAGGACTCTCTGCTTCCCATCTTTCAGCGAATGC 1331

Qy 1221 gaggctgcggcgagaaagaaacccgaaggtttccgcgggggagcggggttcgaccccgag 1280

Db 1332 CTTTACTCGCTCTCTCTCCCTCCACTCTGTCCTCTGTCCTCTGCGCTGCCAGAGTTTAAAGACAG 1391

Qy 1281 cctcgcgcgagagagagcccgcggttaacaagcagcgagcggcgcgctcgcgcgc 1340

Db 1392 CAGCAGCTCCCTCACAGATCCCGAGTTTCTTACGTCAAGAGTCCAGCTGACAGAGAGGG 1451

Qy 1341 cacgaggaagagagagag---tcgttaagatagagaccagcagagataaacctcagcc 1397

Db 1452 CAGTGAGAGAGTGGCTGGAGGCCCGCCGACACACCATCGGCCACGCCCTCGGGCCCCCGC 1511

Qy 1398 ccgcaagcacaacagagctccctgagctgagctcctcacccttccgagctgtctgtgaa 1457



Db 1512 ACTCCAGCTGGTCCAGGCGCTTTCCACGAAGAGTGCTCTCAGCCTTCTCCCTCCAGCGGAA 1571  
Qy 1458 ctgcaggggggagtgccctgccaacgtgacctgtctctccaaacctgagccccgacac 1517  
Db 1572 GAGCCCCCTGAGTCTGAGCCCTGCATTTACAGGTGCCCTT-----TGGGCCACAGCTGA 1622  
Qy 1518 caagcagggcctctcccttgatcagccgcgtgctgaaagaccagcgctgccccggagcga 1577  
Db 1623 ACAGGGCTCATCCCACTCTGAGCCCTCTCCTCAGTATGCTGGCGAGCGCGGATGGA 1682  
Qy 1578 cgatgagatgagggccggaggagagatctcccaactgataaagcgtactctcatagctaa 1637  
Db 1683 CTGCGAG---GAGCCCAACACAAACGCGCTGCCAGACAGAGGCCCTACTTCATAGTCAA 1739  
Qy 1638 ggaagtctaccacgagcgaacatactgaaggtatctgaaggtatctgaaggtatctggtt 1697  
Db 1740 AGAGTTCTCGCTACAGAACCAACATACCTCAAGGATTTAGAAATTTATACCTGTGTGTT 1799  
Qy 1698 tcagagcacagtgcgaagagcgcctgcccgaagcactgaaaagtctcatattccc 1757  
Db 1800 CCGCAGCGCAGTGGTGAAGGAGGACGCCATGCTCGCACTCTGATGACGCTGCTCTCTC 1859  
Qy 1758 gaattttgaacctttgcacaaatttcataactaatcttctcaagaaattgagcaagact 1817  
Db 1860 CAACATCGATCCCATCTATGAGTTCCACAGAGGCTTCTCGCGGAGGTGGAGCAGAGGCT 1919  
Qy 1818 tgcctgtggaagggcgctcaaatgcccgaatacagag---attacaaagaatcgcgga 1874  
Db 1920 GGCACCTGGGAGGGGCCCTTCAAGGCCACCAAAAGGAGCGATCATCAAGAAATCGGGGA 1979  
Qy 1875 tgcctatgaaagaacattcaggcgtgagcaccctggcggtcacctctggaagcacag 1934  
Db 1980 CATCTCTCAGGACATGCGCCAGTTAAAGGAGTTTACCAGCTACTTCCAAACACATGA 2039  
Qy 1935 cgaggccttgaggccctggagaatggaatcaagagctcccgcggtggagaaactctg 1994  
Db 2040 CGAGGCTCTAACGAACCTGGAAGGCTTACCAACGCTGTAAAGATTTGGAGGAGGTGA 2099  
Qy 1995 cagagacttgagctcagagagtggttacctacgctcaacacctctctctgggcc 2054  
Db 2100 CAAGGATTTGAGCTGAGAAGGTCTGCTACTTGCCTCTCAACACGTTCTGCTGAAGCC 2159  
Qy 2055 actgcacggctcatgcactacaagcaggtccctggagcggctgtgcaaacaccaccgcc 2114  
Db 2160 CATCCAGCGCTGCTGCACATACCGCTGCTGCTGCGCGCTATGCGGACATATACAGCCC 2219  
Qy 2115 gagccacgcgaacttcaggactgcgagccgcttttgacagagatacagagatggtg 2174  
Db 2220 CGGGCACCATGACTACGCTGACTGCCATGACGCCCTTGAAGGCCATCACAGAGGTGACCAC 2279  
Qy 2175 acagctccacggtacgatgatcaagatggagaaatttccagaagctgcagaaactcaaga 2234  
Db 2280 CACACTACACACATCTCATCGGCTGGAGAACCTGCAAGACCTAACGGAGTGCAGCG 2339  
Qy 2235 agatttgattggcattgacaatctgtggttcggggaagaggttcatcctgctgggcag 2294  
Db 2340 GGACCTGGTGGCATAGAGAACCTCATTTGCTCTGCGCAGGAGTTTCATCGTGAGGCGTG 2399  
Qy 2295 cctcagaacactctcggggaagggctccagcagcagcatgtctctctctgctcaacagct 2354  
Db 2400 CTTTCACAACTCACCAGAAGGCGCTGCGACAGAGGATGTTTTTCTGTTCTCAGATAT 2459  
Qy 2355 cctgctatacacagcggggtgcagcgctccaatcagtttaaaagtccacggcgagct 2414  
Db 2460 GTTGCTGTACACAAAGCAGAGTTGCAGGGACACGCCACTTCGCGATCCGGGCGCTCTCT 2519  
Qy 2415 cccgctctatggtgactgacgattgagagagcgaagcagtgagggtgggggtccccactgcct 2474  
Db 2520 TCCCTCCAAAGGCATGCTGTGTGAAGAAAGTATACAGAGTGTGCTGTTCCACACTGTTT 2579  
Qy 2475 gaccttcggggccagcggcgagtcacatcgtggtggccgagctctctggtccagatgga 2534  
Db 2580 CACCATCTACGCGCTCAGAAAAAATCGTGGTGGCAGCCAGCAGCTCGGCTGGAGAAGA 2639

Qy 2535 gaagtgggttgagacatccagatgcccattgacctggcgagagaagagcagcagccccgc 2594  
Db 2640 GAAGTGGATGCTGACCTGAACCTCCGCTGAACCTCCGAGTCCCAAGCAGCAAGAGTGGGTGACAGCGC 2699  
Qy 2595 ccttgagttctctggccagca-----gcccccctgacaacaagtccccctctgatgaagccac 2648  
Db 2700 CCTTGCACTGCCAGGCGGCACTGTGTGCACTGCTCCCGCCAGATCCCGCCCAACAGGATATC 2759  
Qy 2649 cgcggttgaccagagtgatcagagtgacctgagcgctcgcgcacatcgcgtgagcgcca 2708  
Db 2760 TCTG---GAGCAGAGAGTCAGAAGATGATGCTCGGGGTGTCCGAGCTCCCTGGAGGGGCA 2816  
Qy 2709 gccccgcacccgcggcaacaacaatggtgcacgtgtgctggcagcgcgaacacccagcgtctc 2768  
Db 2817 TGGCCAGCAGCCGGGCCAACACCAACATGACAGGTGTGTGTACCCGGAACACACGCGTGC 2876  
Qy 2769 catggtggaacttcagatcgcagtgagaaatcagtgctggaacacctgctgaggaatt 2828  
Db 2877 CAGGCGAGACACAGTGCACCTGTCGAGAACACAGCTTTCAGGATATCTGCTAAGAAAGTT 2936  
Qy 2829 caaaacagcaacgggtggcagaagctgtggtgggtgttcacaaactctgcctgttctt 2888  
Db 2937 CAAAAACAGCTCATGCTGGTGGCAGAAGCTCTGGTCTCTTTACCAACCTCTCTGTTTCTT 2996  
Qy 2889 ctacaaatcacacacagacaatcatcccttgcagcctgcctctgcctactcgt 2948  
Db 2997 CTAAAAATCATCAGAGTACTACCCACTGGCCAGCTCCCGCTGCTGGGTACACGCT 3056  
Qy 2949 caccatccctctctgagtcgcgagaacatccagaagactcagtggttcaagctgcactcaa 3008  
Db 3057 GAGCATCCCCAGGAGGCGCGATGSCATACACAAAGACTATGTTTCAAGCTCCAGTTCAA 3116  
Qy 3009 gtccacgtctactacttcaggcggaagcagagtgacacgttgcgaaggtcggaagt 3068  
Db 3117 ATCCACGCTACTTCTTCGCGCTGAGAGCAAGTACACATTTGAAAGGTGGATGGAGT 3176  
Qy 3069 gatccgagtcgaccagctctcctcgacccc 3103  
Db 3177 GATCCAGGGGGCCAGCAGCTCAGCCGGGAGGGCCC 3211

## RESULT 4

BC009153 3719 bp mRNA ROD 12-JUL-2001  
LOCUS Mus musculus, Similar to KIAA0793 gene product, clone MGC:6304  
DEFINITION IMAGE:2855209, mRNA, complete cds.  
ACCESSION BC009153  
VERSION BC009153.1 GI:14318718  
KEYWORDS MGC.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 3719)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (05-JUN-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgabs-r@mail.nih.gov](mailto:cgabs-r@mail.nih.gov)  
Tissue Procurement: Gilbert Smith, Ph.D.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Baylor College of Medicine Human Genome  
Sequencing Center  
Center code: BCM-HGSC  
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>  
Contact: [villalon@bcm.tmc.edu](mailto:villalon@bcm.tmc.edu)  
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia,



A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,  
Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAK plate: 5 Row: m Column: 20.

## FEATURES

Location/Qualifiers

Source

1..3719  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/clone="MGC:6304 IMAGE:2655209"  
/tissue\_type="Mammary tumor. MMTV-LTR/INT3 model. 5 month  
old mouse. Taken by biopsy."  
/clone\_lib="NCI\_CGAP\_Mam2"  
/lab\_host="DH10B"  
/note="Vector: pCMV-SPORT6"  
99..3296

## CDS

/codon\_start=1  
/product="Similar to KIAA0793 gene product"  
/protein\_id="AAH09153.1"  
/db\_xref="GI:14318719"  
/translation="MGEIEGYRALPTSGRLGGQTAIGVSTLEPQSLSPRMQEKHM  
RIRVLDSSTNVEFLRPEKDCQVLLTQWQHLNLECDYFGLFKNVSYWIMLEPM  
KPIROVKPKNAVLRVAKFPDPGLOEYTRYLFALQIKRDLERLTCTANTA  
ALLISHLQSEIGDYDEFLDRHLKANEYLNQESLLEKILDFHQHRTGQPAESDFQ  
VLEIARKLEMGIIRFHMSADREGTKINLAVSHMGLVFGQTKINFNWSRVKRLUSFK  
RKRFLIKLHPVHGYPQDTLEFLFGSDECKNFWKICVEYHTFFRLSDQPKAKAFV  
FSRSFSYSGRTQQLVDYVKGDMKRPYERRHSKRTSLHALTVLDPKPSVSFTD  
GLRTSASISANTSPYPPSPGLPNLKDSSSLVDPQAPVIKGTAABRSSGPD  
SSDGPSTOAHLPGLPVRPGFSDPSQPSLSKSHLSLCELOALSTPAEOGAS  
PVLSVPLSGATARDNDSEQKHMPDEAFYIAKEILATERTYLLKLEVITYWFRS  
VLKEEAPMALALFLEFNIDPIEFHFGFLHEVQRLAMEGSPSAHLKHQDHQIGD  
ILLNRMLQRTSYFQRHDEYLELEKATCKKLEAVYEFELQKVCYLPINFTLL  
KPVQRLYRLLSLCAHYSPGHRDYADHEALKAITVETVLOQSILTRLENLQKLT  
ELQDLGVENLIRGREFIRGCLHKLTKKGLQRMFFLFSDMLLYTSKSVTGASHE  
RIRGLPLRGMLVESENEMSVLHCFITYAAKTIIVAASTLEKEMQDLNLAIOA  
AKTGDSPVLLGGVYVYTRPSDEVSLESEDEGRNGRSLGNSQHRANTTMVWC  
YNTSVSRADSAVENSGLTLLRFRKNSNGKWLWVTFNFCFLFFYKTHODDYLPLA  
SLPLGYSVLPREADSHTKDYFLKQKSHVYFFRAESKYTFERWMDVIKRASSSPG  
RPPSTQDCSHHSGLEAIREKEACPSPCLDKNL"

BASE COUNT 997 a 979 c 929 g 814 t

## ORIGIN

Query Match 32.3%; Score 1013.2; DB 10; Length 3719;  
Best Local Similarity 59.3%; Pred. No. 7.3e-216;  
Matches 1846; Conservative 0; Mismatches 1238; Indels 30; Gaps 6;  
QY 22 ccgacccagatcacgaactggggcccgaaatctgggagatcagtaacctggaaagt 81  
DB 132 CCAACTTCAGGACACGCTGGGTGGCCAGACCGCCCAATTGGAGTGAGTACCTCGGAGCCA 191  
QY 82 ggaagaagccgccccacacacttcaggaaactcgtgtccataaaatccagatgctg 141  
DB 192 GACGACGCTGTACCCAGGATGAGGAGAGACACATCGCATCAGATGAGCTGCTG 251  
QY 142 gatgacaccagaggcattgaagtccacaaagagctcctgggaaggtgctgctggat 201  
DB 252 GACGACGCTGTAGAGTATTTCATCGACCGCAAGTGTGACGGGACGCTTACTGACT 311  
QY 202 gcagtttgaaacacactcaactcgtggaggtgactattttggcctcagatttctgat 261  
DB 312 CAAGTGTGAAGCAATTTGAATTCATTTGAATGTGACTACTTTGGTGGAGTTCAAGAA 371  
QY 262 cacaaaaagatcacggtggtcgtatcctctaaacccatgtgaacagattagaag 321  
DB 372 GTCCACTCTACTGATTTGGCTTGAACTATGAAACCCATCATNATAGGCAAGTACGAAG 431  
QY 322 ccaaaagcaggtgtgttaagtttggtgaaattcttccgctgacacacacacac 381  
DB 432 CCAAAAAATGCGGTGCTTCGCTGGCAGTAAATTTTCCCGCCTGACCTGCTGACGTC 491  
QY 382 caagaagaactcacaaggtaacctgttctcgctgcaggtgaagcaggaacttgctcaaggc 441



```
/dev_stage="adult"
/tissue_type="melanoma (MeWo cell line)"
polyA_signal 2852..2857
polyA_site 2859
BASE COUNT 879 a 592 c 620 g 797 t
ORIGIN

Query Match
Best Local Similarity 10.9%; Score 342.8; DB 9; Length 2888;
Matches 476; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 397 aggtacctgttcgctgcagtgagcaggacttgctcaaggagggtgacgtgtaata 456
Db 123 AGGTATCTTTTACTCTCTCAATAAAGAGGATTTGGCTCTAGGAAGGCTTCATGCA 182
QY 457 gacacagcgagctctctgtttcacacattgccaatctgcaatctgagattggtttat 516
Db 183 GACACTGTACAGCTGTGATGATCTACATCTTACATATCAGACTTGGAGCTTTCAT 242
QY 517 gaagccttgacagagacacttagcaaaaaataataacataccctcagcaagacaccta 576
Db 243 GAAGAACAAGATAGGAAGCATCTGGCACAACTCGTACTTACCAAAACCAAGACTGTTA 302
QY 577 gaggcaaaatcgtggaatttcaccataaaccacattggacaaaccaccagacaatcagat 636
Db 303 GAGGCAAGATCATGCATCTTTCATCAGAAGCAGCATTTGGCAGAGGCCAGCTGAATCTGAC 362
QY 637 tccagctctagagatgcccgtgcgttagatgtagatgtagatgtagatgtagatgtagat 696
Db 363 ATCTGCTACTGGACATAGCAAGAGCTGGATATGATGGCATCAGGCTCACCCGCC 422
QY 697 aaggcagggagggacagaaatcaatctgcccgttgccaaacacagggaaattctagtgtt 756
Db 423 ACTGATGGTGAAGGATGCAGATTACCTCGCTGTGCTTCACATGGAGTACTGTGTTA 482
QY 757 caaggtttcaataataatccttcaactcagggcagggcagggcagggcagggcagggcag 816
Db 483 CGGGGAAATACAAAGATCAATCTTTAACTGGGTAATAATCCGAAGTTGAGTTTAA 542
QY 817 aggaagcgtttctcatcaagctccgcccagatgccaatagtcgctaccagatcaccttg 876
Db 543 AGAAGCATTTTCTCATCAAACTTCATGCCAATATCTTGGTGTGTCAGGATACCTTG 602
QY 877 gaattcctgatggcagtcgtggttcttgcgaagtccttctggaataatctgtgtgaaacat 936
Db 603 GAGTTCCAGTTCGCTATAGTGGACGAACCCAAAGCACTTTTGGATATGGGAGA 782
QY 937 catgccttcttagacttttgaagagcccaaaacaaagcccaagcccgctcctctttagc 996
Db 663 CATGCTTTCTCAGGCTTTCGGAAGAGCCCAATCAAAAGCCCAAAACCTACTCTGCAGC 722
QY 997 cgggggtcaccatttcggttcagtggtcgactcagaagcaggttctgactatgttaa 1056
Db 723 AAGGTTCCAGTTTCGCTATAGTGGACGAACCCAAAGCACTTTTGGATATGGGAGA 782
QY 1057 gaaggaggacataaagagtgtagtttgaaggaagca 1094
Db 783 AAAGGAGGCTGAAGAGCTTGCCATTGAAAGGAAACA 820

RESULT 6
HSM801276 1438 bp mRNA PRI 18-FEB-2000
LOCUS Homo sapiens mRNA; cDNA DKFp434M2221 (from clone DKFp434M2221);
DEFINITION partial cds.
ACCESSION AL122052
VERSION AL122052.1 GI:6093247
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```



repeat\_region 18510..18797  
/note="AluJo repeat: matches 1. .302 of consensus"  
repeat\_region 18800..18956  
/note="MLTIA2 repeat: matches 1. .171 of consensus"  
repeat\_region 19392..19704  
/note="AluJo repeat: matches 1. .312 of consensus"  
repeat\_region 19780..20069  
/note="Alu repeat: matches 1. .290 of consensus"  
repeat\_region 22285..22456  
/note="MIR repeat: matches 23. .220 of consensus"  
repeat\_region 22457..22760  
/note="AluSg repeat: matches 1. .304 of consensus"  
repeat\_region 22761..22773  
/note="MIR repeat: matches 220. .230 of consensus"  
repeat\_region 26734..26961  
/note="L1MB2 repeat: matches 5944. .6170 of consensus"  
repeat\_region 28074..28174  
/note="MER44C repeat: matches 7. .103 of consensus"  
repeat\_region 28175..28840  
/note="MER72 repeat: matches 5. .668 of consensus"  
repeat\_region 28841..29431  
/note="MER44C repeat: matches 103. .721 of consensus"  
repeat\_region 29821..30105  
/note="AluSx repeat: matches 1. .285 of consensus"  
repeat\_region 30219..30775  
/note="L2 repeat: matches 2150. .2749 of consensus"  
repeat\_region 30924..31207  
/note="AluSx repeat: matches 3. .294 of consensus"  
repeat\_region 32053..32298  
/note="MIR repeat: matches 1. .257 of consensus"  
repeat\_region 32498..32740  
/note="L1MB3 repeat: matches 5943. .6182 of consensus"  
repeat\_region 33867..33920  
/note="27 copies 2 mer ca 90% conserved"  
repeat\_region 33871..33918  
/note="8 copies 6 mer cacaca 93% conserved"  
repeat\_region 33878..33921  
/note="11 copies 4 mer acac 97% conserved"  
repeat\_region 34367..34420  
/note="9 copies 6 mer ctcttc 75% conserved"  
repeat\_region 34458..34576  
/note="MIR repeat: matches 94. .219 of consensus"  
repeat\_region 34914..35006  
/note="MIR repeat: matches 150. .242 of consensus"  
repeat\_region 35044..35414  
/note="MLT1B repeat: matches 1. .386 of consensus"  
repeat\_region 35440..35480  
/note="L2 repeat: matches 2423. .2464 of consensus"  
repeat\_region 35653..36028  
/note="THE1B repeat: matches 1. .359 of consensus"  
repeat\_region 36683..37012  
/note="AluSx repeat: matches 1. .308 of consensus"  
repeat\_region 37526..37818  
/note="AluJo repeat: matches 1. .290 of consensus"  
repeat\_region 37957..38028  
/note="36 copies 2 mer tg 86% conserved"  
repeat\_region 37968..38031  
/note="16 copies 4 mer gtgt 90% conserved"  
repeat\_region 37980..38027  
/note="8 copies 6 mer gtgtgt 100% conserved"  
repeat\_region 38553..39205  
/note="L1MB3 repeat: matches 5545. .6183 of consensus"  
repeat\_region 39561..39944  
/note="MSTA repeat: matches 1. .426 of consensus"  
repeat\_region 40827..41086  
/note="AluSx repeat: matches 1. .259 of consensus"  
repeat\_region 41793..42089  
/note="AluSg repeat: matches 1. .297 of consensus"  
repeat\_region 42330..42393  
/note="L2 repeat: matches 2627. .2701 of consensus"  
repeat\_region 42692..42857  
/note="L1MB2 repeat: matches 5654. .5816 of consensus"  
repeat\_region 42858..43151

/note="AluSg repeat: matches 1. .293 of consensus"  
43152..43183  
/note="L1MC2 repeat: matches 5816. .5846 of consensus"  
repeat\_region 43184..43479  
Query Match 8.1%; Score 254; DB 9; Length 96183;  
Best Local Similarity 100.0%; Pred. No. 2.3e-46;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1162 cagctcagcagcagcagcagccttacatttgagagagtgccgaatctctcagggggccag 1221  
|||||  
Db 95665 CAGTCTCAGCAGCAGCAGCAGCCTTACATTGGAGAAGGTGCCGAATCTCCAGGGGGCCAG 95724  
|||||  
QY 1222 agctccggcgaggaagaaacgaaggtttccgcgggagccgggtcgccaccgagc 1281  
|||||  
Db 95725 AGCTCCGCGCAGGAGAAAGAACCGAAGGTTTCCGCCGGGAGCCGGGTGCGACCCGAGC 95784  
|||||  
QY 1282 cctgcgcgcagagagagcccccgcgggtaacagcagcgagcgagcgccctcgcgccc 1341  
|||||  
Db 95785 CTTGCCCGCAGGAGAGAGCCCCCGGGTAAACAAGCAGCGAGCGAGCCCTCTCGGGGCC 95844  
|||||  
QY 1342 acggaggaagagagaggtcgttaagatagagccagcagagtaaacctcagccccc 1401  
|||||  
Db 95845 ACGGAGGAGAGGAGGAGGTCTTAAGGATAGGACCCAGCAGAGTAAACCTCAGCCCCCG 95904  
|||||  
QY 1402 cagccaagcacagg 1415  
|||||  
Db 95905 CAGCCAAGCACAGG 95918  
|||||  
RESULT 8  
AC022669/c  
LOCUS  
DEFINITION Homo sapiens chromosome 15 clone RP11-96B23 map 15, WORKING DRAFT  
SEQUENCE, 33 unordered pieces.  
ACCESSION AC022669  
VERSION AC022669.4 GI:9966291  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 152053)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 15, clone RP11-96B23  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 152053)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,  
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,  
Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,  
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,  
Ferrelira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,  
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,  
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,  
MacDonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,  
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,  
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,  
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,  
Stojanovic,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,  
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,  
Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Sep 4, 2000 this sequence version replaced gi:7229803.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- Project Information  
Center project name: I4938  
Center clone name: 96\_P\_23  
----- Summary Statistics  
Sequencing vector: M13; M77815; 93% of reads  
Sequencing vector: Plasmid; n/a; %0.f% of reads  
7.39091718610864Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 136451 bases at least Q40  
Consensus quality: 143612 bases at least Q30  
Consensus quality: 146365 bases at least Q20  
Insert size: 158000; agarose-fp  
Insert quality: 148853; sum-of-contigs  
Quality coverage: 3.4 in Q20 bases; agarose-fp  
Quality cov.  
NOTE: This is a 'working draft' sequence. It currently consists of 33 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 784: contig of 784 bp in length  
785 884: gap of 100 bp  
885 1956: contig of 1072 bp in length  
1957 2056: gap of 100 bp  
2057 3113: contig of 1057 bp in length  
3114 3213: gap of 100 bp  
3214 4817: contig of 1604 bp in length  
4818 4917: gap of 100 bp  
4918 6582: contig of 1665 bp in length  
6583 6682: gap of 100 bp  
6683 8059: contig of 1377 bp in length  
8060 8159: gap of 100 bp  
8160 22369: contig of 14210 bp in length  
22370 22469: gap of 100 bp  
22470 24974: contig of 2505 bp in length  
24975 25074: gap of 100 bp  
25075 27206: contig of 2132 bp in length  
27207 27306: gap of 100 bp  
27307 29420: contig of 2114 bp in length  
29421 29520: gap of 100 bp  
29521 32000: contig of 2480 bp in length  
32001 32100: gap of 100 bp  
32101 36330: contig of 4230 bp in length  
36331 36430: gap of 100 bp  
36431 39067: contig of 2637 bp in length  
39068 39167: gap of 100 bp  
39168 42564: contig of 3397 bp in length  
42565 42664: gap of 100 bp  
42665 46664: contig of 4000 bp in length  
46665 46764: gap of 100 bp  
46765 51047: contig of 4283 bp in length  
51048 51147: gap of 100 bp  
51148 55269: contig of 4122 bp in length  
55270 55369: gap of 100 bp  
55370 58831: contig of 3462 bp in length  
58832 58931: gap of 100 bp  
58932 64077: contig of 5146 bp in length  
64078 64177: gap of 100 bp  
64178 68808: contig of 4631 bp in length  
68809 68908: gap of 100 bp  
68909 74170: contig of 5262 bp in length  
74171 74270: gap of 100 bp  
74271 79296: contig of 5026 bp in length  
79297 79396: gap of 100 bp  
79397 84724: contig of 5328 bp in length  
84725 84824: gap of 100 bp

\* 84825 89713: contig of 4889 bp in length  
89714 89813: gap of 100 bp  
89814 94721: contig of 4908 bp in length  
94722 94821: gap of 100 bp  
94822 100196: contig of 5375 bp in length  
100197 100296: gap of 100 bp  
100297 107015: contig of 6719 bp in length  
107016 107115: gap of 100 bp  
107116 114507: contig of 7392 bp in length  
114508 114607: gap of 100 bp  
114608 122671: contig of 8064 bp in length  
122672 122771: gap of 100 bp  
122772 128052: contig of 5281 bp in length  
128053 128152: gap of 100 bp  
128153 138583: contig of 10431 bp in length  
138584 138683: gap of 100 bp  
138684 150248: contig of 11565 bp in length  
150249 150348: gap of 100 bp  
150349 152053: contig of 1705 bp in length.  
FEATURES  
Location/Qualifiers  
1. .152053  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="15"  
/map="15"  
/clone="RP11-96B23"  
/clone\_lib="RPC1-11 Human Male BAC"  
1. .784  
/note="assembly\_fragment"  
vector\_end:SP6  
vector\_side:left  
885. .1956  
/note="assembly\_fragment"  
2057. .3113  
/note="assembly\_fragment"  
3214. .4817  
/note="assembly\_fragment"  
4918. .6582  
/note="assembly\_fragment"  
6683. .8059  
/note="assembly\_fragment"  
8160. .22369  
/note="assembly\_fragment"  
22470. .24974  
/note="assembly\_fragment"  
25075. .27206  
/note="assembly\_fragment"  
27307. .29420  
/note="assembly\_fragment"  
29521. .32000  
/note="assembly\_fragment"  
32101. .36330  
/note="assembly\_fragment"  
36431. .39067  
/note="assembly\_fragment"  
39168. .42564  
/note="assembly\_fragment"  
42665. .46664  
/note="assembly\_fragment"  
46765. .51047  
/note="assembly\_fragment"  
51148. .55269  
/note="assembly\_fragment"  
55370. .58831  
/note="assembly\_fragment"  
58932. .64077  
/note="assembly\_fragment"  
64178. .68808  
/note="assembly\_fragment"  
68909. .74170  
/note="assembly\_fragment"  
74271. .79296  
/note="assembly\_fragment"

misc_feature	79397..84724	/note="assembly_fragment"
misc_feature	84825..89713	/note="assembly_fragment"
misc_feature	89814..94721	/note="assembly_fragment"
misc_feature	94822..100196	/note="assembly_fragment"
misc_feature	100297..107015	/note="assembly_fragment"
misc_feature	107116..114507	/note="assembly_fragment"
misc_feature	114608..122671	

Query Match 8.1%; Score 254; DB 2; Length 152053;  
Best Local Similarity 100.0%; Pred. No. 2.2e-46;  
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1162	cagtcacgacgagcaccgccttacatttgagaagggtgccgaatctccagggggccag	1221
Db	108632	CAGTCTCAGCAGACACCAGCGCTTACATTTGGAGAAGGTGCCGAATCTCCAGGGGCGCAG	108573
QY	1222	agctcgccgcgagaaagcaagcaagtgttccgccggggagcggggtcgcaccccagc	1281
Db	108572	AGCTGCCGGCGAGAAAGAACCAAGGTTTCGCCGGGGAGCGGGGTGCACCCCAGC	108513
QY	1282	cctgcgcgaggaagaaqccccgcgggttaacaagcaggcgagcgagcgctcgcgccc	1341
Db	108512	CTGCGCCGAGGAGAGCGCCCGGGGTAAACAAGCAGCGGAGCGCGCTCTCGGCGCCC	108453
QY	1342	acggaggaagaggagggtcgtttaagatatgagaccacgacagadtaaactcagccccgc	1401
Db	108452	ACGGAGGAAGAGGAGGAGGTGCTTAAGGATAGGACCAGCACAGACTAAACCTCAGCCCCG	108393
QY	1402	cagccaagcacagg	1415
Db	108392	CAGCCAAGCAGCAGG	108379

RESULT 9

AL137249

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

AL137249 181800 bp DNA HTG 03-AUG-2001  
Homo sapiens chromosome 13 clone RP11-111L24 map q31.3-32.3, \*\*\*  
SEQUENCING IN PROGRESS \*\*\*, 4 unordered pieces.  
AL137249 GI:15131444  
HTG; HTGS\_PHASE1; HTGS\_ACTIVEFIN; HTGS\_DRAFT; HTGS\_FULLTOP.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 181800)  
Garner,P.  
Direct Submission  
Submitted (02-AUG-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
On Aug 9, 2001 this sequence version replaced gi:15131193.  
----- Genome Center  
Center: Sanger Centre  
Center code: SC  
Web site: http://www.sanger.ac.uk  
Contact: humquery@sanger.ac.uk  
----- Project Information  
Center project name: BA31K22  
----- Summary Statistics  
Assembly program: XGAP4; version 4.5  
Sequencing vector: plasmid; L08752; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Consensus quality: 180516 bases at least Q40  
Consensus quality: 180899 bases at least Q30  
Consensus quality: 181168 bases at least Q20  
Insert size: 181500; sum-of-contigs

```

Insert size: 174530; 1.8% error; agarose-fp
Quality coverage: 6.26x in Q20 bases; sum-of-contigs Quality
coverage: 6.72x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 114302: contig of 114302 bp in length
* 114303 114402: gap of 100 bp
* 114403 135098: contig of 20696 bp in length
* 135099 135198: gap of 100 bp
* 135199 173346: contig of 38148 bp in length
* 173347 173446: gap of 100 bp
* 173447 181800: contig of 8354 bp in length.
*
FEATURES
      source
      1. 181800
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="13"
      /map="q31.3-32.3"
      /clone="RP11-111L24"
      /clone_1fb="RPC1-11.1"
      1. 114302
      /note="assembly_fragment:00010"
      fragment_chain:1
      clone_end:SP6
      vector_side:left
      114403..135098
      /note="assembly_fragment:01859"
      fragment_chain:1
      135199..173346
      /note="assembly_fragment:02403"
      fragment_chain:1
      173447..181800
      /note="assembly_fragment:02166"
      fragment_chain:1
      clone_end:r7
      vector_side:right"
BASE COUNT 49780 a 43544 c 41179 g 46996 t 301 others
ORIGIN

Query Match      8.1%; Score 254; DB 2; Length 181800;
Best Local Similarity 100.0%; Pred. No. 2,1e+46;
Matches 254; Conservative 0; Mismatches 0; Indels 0; Gaps

QY 1162 cagctcagcagagaccagcagccttacatttgagaaaggtgcgaatctccaggggccag 1221
|||||
Db 36237 CAGGTCACGACAGACACAGCCTTACATTTTCGAGAAGGTGCCGAATCTCCAGGGGGCCAG 36296
|||||

QY 1222 agctgcgcgcagagaaagcaaggtttccgcgcggggagccgggtctgcacccgagc 1281
|||||
Db 36297 AGCTGCGCGGCAGGAAGAAACCGAGGTTTCCGCCGGGGAGACCGGGGTTCGCACCCGAGC 36356
|||||

QY 1282 cctgcgcgcagagaaagcccgcgggtaacaagcagcgcgcgcgcgcgcctcgcgcgc 1341
|||||
Db 36357 CCTGCCCGGAGGAAGCCCGCGGGTAAACAAGCAGCGCGGACGCGGCTCTCGCGCGCC 36416
|||||

QY 1342 acgagagaagaggagggttcgttaagatagtagccaccagcagagtaaacctcagccccc 1401
|||||
Db 36417 ACGGAGAAGAGGAGGAGGTGTTTAAAGATAGGACCCAGCAGAGATAAACCTCAGCCCCC 36476
|||||

QY 1402 cagccaagcacagg 1415
|||||
Db 36477 CAGCAACACACAGG 36490
|||||

RESULT 10

```

RESULT 10



G22662/c  
LOCUS human STS 483 bp DNA STS 31-MAY-1996  
DEFINITION human STS WI-14178, sequence tagged site.  
ACCESSION G22662  
VERSION G22662.1 GI:1342988  
KEYWORDS STS; STS sequence; primer; sequence tagged site.  
SOURCE human STSs derived from sequences in dbEST and the Unigene collection.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 483)  
Hudson, T.  
Whitehead Institute/MIT Center for Genome Research; Physically  
Mapped STS  
Unpublished (1995)  
Contact: Thomas Hudson  
Whitehead Institute/MIT Center for Genome Research  
Whitehead Institute for Biomedical Research  
9 Cambridge Center, Cambridge MA 02142 USA  
Tel: 617 252 1900  
Fax: 617 252 1902  
Email: thudson@genome.wi.mit.edu  
Primer A: ATGTTTTCGATTTTCAGACAGGTG  
Primer B: GAGTCTCTTGTGATTGATGGCC  
STS size: 127  
PCR Profile:  
  Presoak:  
  Denaturation:  
  Annealing: 56 degrees C  
  Polymerization:  
  PCR Cycles: 35  
  Thermal Cycler:  
Protocol:  
  Template: 10 ng  
  Primer: each 5 pM  
  dNTPs: each 4 nM  
  Taq Polymerase: 0.025 units/ul  
  Total Vol: 20 ul  
Buffer:  
  MgCl2: 1.5 mM  
  KCl: 50 mM  
  Tris-HCl: 10 mM  
  pH: 9.3  
Derived from dbEST (genbank accession T81406).  
FEATURES  
  source  
    1..483  
      Location/Qualifiers  
        /organism="Homo sapiens"  
        /db\_xref="taxon:9606"  
        /map="291.8 CR from top of Chr13 linkage group"  
        /nap="143  
        17..143  
        17..40  
        complement(121..143)  
        127 t 10 others  
        121 a 103 c 122 g 127 t  
STSS  
  primer\_bind 17..40  
  primer\_bind complement(121..143)  
BASE COUNT 121 a 103 c 122 g 127 t 10 others  
ORIGIN  
Query Match 6.7%; Score 210; DB 11; Length 483;  
Best Local Similarity 87.8%; Pred. No. 2.4e-36;  
Matches 251; Conservative 0; Mismatches 29; Indels 6; Gaps 2;  
QY 2856 gtgggtgggtgttcacaaactctgctgttcttctcacaat-----cacaccaggagcaatc 2911  
  |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  
Db 414 GTTGGTGTTCANAAATTTCTCGCTGGTCTTTANAAATTTNAAACCCAGGACCAATC 355  
QY 2912 atcccttgcca--gctgcctctgctcggtactcgctcaccatccctctgagtcgga 2969  
  |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  
Db 354 ATCCCTTTCAGACCTGCTCTCTCGGNTACTCGCTCACCATCCCTCTGAGTCCGA 295

QY 2970 gaacatccagaagaactacgtgttcaagctgcacttcaagtcaccagctcactactctcag 3029  
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  
Db 294 GAACATCCAGAAAGACTACGTGTTCAAGCTGTCACCTTCAAGTCCACGCTACTACTTCAG 235  
QY 3030 ggcggaagcgaagtagacacgttccgaaagggtgagatgggaagtgatccgcagtgccaccagctc 3089  
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  
Db 234 GCGGAAAGCGAGTAGACACGTTCCGAAAGGTGGATGGAGTAGTCCGCGAGTCCACAGCTC 175  
QY 3090 tgctcgacacccacgctgttgagccacaaagagctctctgtgtat 3135  
  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  
Db 174 TGCTCGGACCCACGCTGTTGAGCCACAAAGAGTCTCTTGTGTAT 129  
RESULT 11  
AF156225  
LOCUS Homo sapiens 2595 bp mRNA PRI 04-JUL-1999  
DEFINITION Homo sapiens protein 4.1 (EPB4.1) mRNA, complete cds.  
ACCESSION AF156225  
VERSION AF156225.1 GI:5353737  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
  1 (bases 1 to 2595)  
  AUTHORS Huang, S.C., Wang, C., Lichtenauer, U., Vortmeyer, A. and Zhuang, Z.  
  TITLE Sequence of protein 4.1 from a human neuroblastoma cell line: LAN5  
  JOURNAL Unpublished  
REFERENCE  
  2 (bases 1 to 2595)  
  AUTHORS Huang, S.C., Wang, C., Lichtenauer, U., Vortmeyer, A. and Zhuang, Z.  
  TITLE Direct Submission  
  JOURNAL Submitted (03-JUN-1999) NCI, NIH, 9000 Rockville Pike, Bethesda, MD 20892, USA  
FEATURES  
  Location/Qualifiers  
    1..2595  
      /organism="Homo sapiens"  
      /db\_xref="taxon:9606"  
      /chromosome="1"  
      /map="1p36"  
      /cell\_line="LAN5"  
      /cell\_type="neuroblastoma"  
      1..2595  
      /feature="EPB4.1"  
      1..2595  
      /feature="EPB4.1"  
      /codon\_start=1  
      /product="protein 4.1"  
      /protein\_id="A42222.1"  
      /db\_xref="GI:5353738"  
      /translation="MTTEKSLVTEAENSQHQKKEGEAINSQGOEQEPOQEEECOTAAE  
      GDNWCEHLKASNGDTPHEDLTNKNRSTENRGLSRFLSPSPKRPKSOVSEEGKEV  
      ESDKKEGGGQKEIEFGTSLDEIILKAPIAAPEPELKTDPDPSLDLHLSLSAETQPAOE  
      ELRDPDFEIKEGEGLEECSTIEVKEESQSKAETELKASQKPIKPHRMHCKVSLID  
      DTVYECVVEKHAQGLLRVCEHLNLEEDYFGLAINDNTTSKTLWLSAKIKKQVR  
      GVPNFTNVKFPDPAQLTDTITRYLCQLRQDINAGRLVAGRLPCSFATLALGSTIQ  
      SELGIDYDPELHGVDYVDFKLPAPNTQKLEEKVMELHKSYSRMTQAQADLELENKK  
      LSMGVDLHKAKDLQGVDIILGVCSSGLLVYKDKLRINRFPMPKVLKISYKRSFPFIK  
      IRPGEQYESTIGFLPSRAAKKLVKVCVEHHTFRLTSTDTIPKSKFLALGSKFR  
      YSGRTQATRQASALIDPAPHFERTASKRASRLDGRAAVDSADSPRPTSPALITQ  
      GVAAEGVLDASAKTIVYPAQKQETVKAEVKKEDEPPEQAEPEPEAKVKEHLEVT  
      VPTSGDQTKLAETEDLIRMRKKRLDGEIYIRHSLMLDELKQSQEIEKHH  
      ASISELKNNEVPEPRPESWDRKRLSTHPTLNINGQIPTGEGPLVKTQVTIS  
      DNANAVKSEIPTKQVPIVHTETKTITYEAAOTDNDNSGDLDPGVLLTAOTITSETPSST  
      TTTQITKTVKGGISETRIEKRIVITGDADIDHDQVLVQAIKEAKEQHPDMSYTKVVVH  
      QETETADE"  
BASE COUNT 881 a 563 c 601 g 550 t  
ORIGIN  
Query Match 6.2%; Score 194; DB 9; Length 2595;  
Best Local Similarity 52.9%; Pred. No. 7.9e-33;  
Matches 491; Conservative 0; Mismatches 425; Indels 12; Gaps 3;







DEFINITION Human structural protein 4.1 mRNA, complete cds.  
ACCESSION M14993  
VERSION M14993.1 GI:182075  
KEYWORDS Human reticulocyte, cDNA to mRNA, clones lambda-HE 4.1-8, lambda-HE  
SOURCE 4.1-6, and lambda-HE 4.1-A.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2867)  
AUTHORS Conboy, J., Kan, Y.W., Shohet, S.B. and Mohandas, N.  
TITLE Molecular cloning of protein 4.1, a major structural element of the  
human erythrocyte membrane skeleton  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83, 9512-9516 (1986)  
MEDLINE 87092279  
REFERENCE 2 (sites)  
AUTHORS Conboy, J., Marchesi, S., Kim, R., Agre, P.P.A., Kan, Y.W. and  
Mohandas, N.  
TITLE Molecular analysis of insertion/deletion mutations in protein 4.1  
in elliptocytosis; Determination of Molecular genetic origins of  
rearrangements  
JOURNAL J. Clin. Invest. 86, 524-530 (1990)  
MEDLINE 90347048  
COMMENT [2] sites: insertions and duplications causing elliptocytosis.  
FEATURES  
source Location/Qualifiers  
1..2867  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
799..2565  
/note="4.1 protein"  
/codon\_start=1  
/protein\_id="AAA35795.1"  
/db\_xref="GI:182076"  
/translation="MHCKVSLDDTVYECVVERHAKGODLLKRVCEHLNLEEDYFGL  
AIVNATSKVWLDAKETIKOVGPWNFTNVPYPPDPAQLTIDITRYVLCQLRQ  
DIVAGRLPCSPATLALGYSYIQSELGDYDPELHGVYVDFKLPAPNOTKELEKVM  
LHKYSRMTQAQDLERLAKLSMYGVDLHAKDLEGVDIILGVSSGLLYVKDL  
RINFPPKVLTKSKSFYIKRPGEEQESTIGFKLPSYRAAKLWKVCVEHHT  
DGAADVSDRSPRTAPAITQGVAEQGVLDASAKTVVPAKQKTVKAEVKKDE  
PPEQAEPEPTAKKRERLDGENIYIRHSLMLEDLDKSOEETKKHHSISELKNE  
MESVPERPSPMDKRLTSHSPFTLNINGQIPTGEPPLKTOFTVTSNANVKS  
PTKDVPIVHETKTIYENAGTVKGGISETRIEKRVITGDADIDHQLVQALKEAK  
EQHPDMSVTKVNVQETEAIDE"  
variation 2016..2017  
/note="duplication insertion of nucleotides 2217-2585  
causing elliptocytosis [2]"  
variation 2017..2256  
/note="deletion causing elliptocytosis [2]"  
BASE COUNT 932 a 620 c 657 g 658 t  
ORIGIN 218 bp upstream of pvuII site.  
Query Match 6.1%; Score 192.4; DB 9; Length 2867;  
Best Local Similarity 52.8%; Pred. No. 1.8e-32;  
Matches 490; Conservative 0; Mismatches 426; Indels 12; Gaps 3;  
Qy 117 cgtgtccatcaaatccagatgctgtagacacacagcaggcattgaagttccacaaag 176  
Db 798 CATGCACCTGCAAGGTTCTTGTGATGACACACTTATGAATGTTGTGGAGAACAA 857  
Qy 177 agctcctgggaagtgctgctggatgacacacacacacacacacacacacacacac 236  
Db 858 TGCTAAGGACCAAGATTTGCTTAACAGATGTGTGAGCATCTCAATCTTTTGGAGAAGA 917  
Qy 237 ctatttggcctcgagtttctctgcacacacacacacacacacacacacacacacac 296  
Db 918 CTAATTTGCTAGCCATTGGGATACACACACCTCTAAGACATGGCTGCGATCCGCCAA 977  
Qy 297 accatttgaaacagattagagcccaagcacgctgttgtagatttggtagaatt 356  
Db 978 AGAATAAAGACAGGTCGTGG---TGTCCTTGGAAATTTACATTAATGTAAGTT 1034

Qy 357 ctttcgcgctgac 416  
Db 1035 TTATCCACTGACCCAGCAGACAGTTAAACAGACACATAAAGATATATATTTATGCTTCA 1094  
Qy 417 ggtgaagcaggactggctcaagcagggttgctgagtgtaataatgacacacgacgactctct 476  
Db 1095 GCTTCGGCAGGACATAGTTGACAGAGCTGCGCTGCTTCTTGTGCAACCTTACCATTTAT 1154  
Qy 477 gatttcacacatttgcaatcgagattgggatttttgatgaagccttgagacagagagca 536  
Db 1155 AGGTTCTTACACCATCCAGTCTGAACCTGGGAGACTACGACCCAGAACCTCATCGGCTGGA 1214  
Qy 537 cttagcaaaaaataataacatacc-----tcagcaagacgacactagagacacaaatcgt 590  
Db 1215 TTATGTTAGTGATTTTAAACTGGCCCGCAATCAGACCAAGGAACCTGGAAGAGGTCAT 1274  
Qy 591 ggaatttcaccataaccacattggcaaacacacacacacacacacacacacacacacac 650  
Db 1275 GGAATGCTATAGTATACAGGTCCATGACTCCAGCTCAGCTCAGGCTGACTTGGAGTTCTTGA 1334  
Qy 651 gattcccgctcggttagatgtatggaaatccggttgccacccgcccagagcagaggaagg 710  
Db 1335 GAATGCCAAAAAGTTGCTCTATGATGGAGTTGATCTTCATAAAGCAAGAGACTTGAAGG 1394  
Qy 711 caggaagatcaatctggcgttgccacacacggaattcttagtggtttcagggtttcactaa 770  
Db 1395 AGTAGATATCATCTCCTAGGTGCTGCTCTAGTGGCTTCTGGTTTACAAAGATAAGCTGAG 1454  
Qy 771 gatcaatgccttcaactggcgaagtgcggaagctgcgagcttcagacacagaggaagcgctttct 830  
Db 1455 AATTAAACCGCTTCCCTTGGCCCAAGAGTGTGAAGATTCTTTATTAACGCTAGTAGCTTTT 1514  
Qy 831 catcaagctccgcgagatgccaatagtcgctaccaggaacaccttgggaattcctgatggc 890  
Db 1515 CATCAAGATTGCGCTGGAGAGCAAGAGCAGTATGAAGTACCATCGGATTCAAACTTCC 1574  
Qy 891 cagtcgggaattcttcaagtcctcttgaaaaatctggtgaacatcatgcctcttttag 950  
Db 1575 CAGTTACCCGAGCAGTAAAGAAATATGGAAGTCTGTGTAGAACATCACACGCTTTTTCAG 1634  
Qy 951 acttttgaagagcaccac 1010  
Db 1635 A---TTGACATCTACAGACACCATTCCTCCCAAGACAAATTTCTTGGCTAGGATCCAAAT 1691  
Qy 1011 tcggttcagtgctcggaactcagaagcag 1038  
Db 1692 TCGATACAGTCCGCGACTCAAGCTCAG 1719  
RESULT 14  
HUMEMP41  
LOCUS Homo sapiens erythroid membrane protein 4.1 mRNA, complete cds.  
DEFINITION HUMEMP41 3043 bp mRNA PRI 03-MAR-1994  
ACCESSION M61733  
VERSION M61733.1 GI:182081  
KEYWORDS erythroid membrane protein 4.1.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 3043)  
AUTHORS Conboy, J.G., Chan, J.Y.C., Chasis, J.A., Kan, Y.W. and Mohandas, N.  
TITLE Tissue- and development-specific alternative RNA splicing regulates  
expression of multiple isoforms of Erythroid membrane protein 4.1  
JOURNAL J. Biol. Chem. 266, 8273-8280 (1991)  
MEDLINE 91217063  
FEATURES Location/Qualifiers  
1..3043  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
816..2741  
/codon\_start=1  
/product="erythroid membrane protein 4.1"



KVEKTHTEVTPTSNGDOTOKLACKGEDLIIRMKKRRERLDGENIYIRHSNLMLEDLD  
KSQEIKKHHASISBLKKNFMSVPEPRPSEMDKRLSHSPFTLNINGQVPTGDGPP  
LVKQTVISDTANAVKSEIPTKDPVIVHTTKTIITYBAAQTESNGDLDPGLVLTAAQ  
TITSSTSTTTTQHTKTVKGGISETRIEKRIIVITGDADIDHDQVLQVAIKEAKEQHP  
DMSVTKVVVHQETEISEE"

exon	2027..2236	
exon	2237..2341	
exon	2342..2384	
exon	2385..2460	
exon	2461..2679	
exon	2680..2767	
exon	2768..2920	
exon	2921..3018	
exon	3019..3191	
exon	3192..3379	
exon	3380..3436	
exon	3437..3478	
exon	3479..3541	
exon	3542..3718	
exon	3719..3847	
exon	3848..3949	
exon	3950..4030	
exon	4031..4135	
3'UTR	4130..5942	
exon	4136..4999	
exon	5000..5942	
BASE COUNT	1640 a 1443 c 1529 g 1330 t	
ORIGIN		

Query Match 6.0%; Score 189.4; DB 10; Length 5942;  
Best Local Similarity 52.5%; Pred. No. 7.8e-32;  
Matches 490; Conservative 0; Mismatches 431; Indels 12; Gaps 3;

Qy	112	aaactcgtgtccatacaaaatccagatgctggatgacacaccaggagggcatttgaagtcca	171
Db	2177	AGAAACATGCATGTAAGGTCCTCTGTTGGATGACACGGTCTACGAATGTGTGTAGAG	2236
Qy	172	caaaagagctctggaaaggtgtgtgtgagtcagtttgcacaccacctcaacctcgtggaa	231
Db	2237	AAACATGCTAAATTTACAGACTTGCTGAAGCGAGTGTGGAGCACCTCAACCTTTTGGAA	2296
Qy	232	ggtgactatttggctcagatttctctgacacaaaagatcacgggtggtggtgatctc	291
Db	2297	GAAGACTACTTTGGTTTAGCCCTGTGGGACAGCGCAACCTCTAAGACATGGCTGGATTCT	2356
Qy	292	ctaaacaccatttgaacacagattagaagccaaagcacgttgttgaagtttgggtg	351
Db	2357	GCCAAAGAAATAAAAAGCAGGTTCGAGG---TGTTCTTGGAAATTCACATTTAATGTG	2413
Qy	352	aaattcttccgctgaccacacacaaactccaagaactcacaaggtacctgttccgcg	411
Db	2414	AAGTTTATCCACCCGACCCAGCAAAATTAACAGAGACATAACAAGATACTATTATGT	2473
Qy	412	ctgacaggtgaagcagagacttggctcaaggcaggttgacgtgtaatgacaccagcagct	471
Db	2474	CTTCAGCTTCGCGCAGGACATTTGCTGGACGCTCTGCCCTGTCTCTTTGCAACTTAGCC	2533
Qy	472	ctcttgatttcacacatigtcaatctgaagattgggatttttgatgaaccttggaacaga	531
Db	2534	CTACTCGGGTCTTACCGATCCGATCTGAGCTGGGAGACTATGACCCCAAGAACTGCACGC	2593
Qy	532	gagcacttagcaaaaaataaaatacat-----acctcagcaagacgacctagagacaaa	585
Db	2594	ATGGATTATGTTAGTGATTTTAACTGGCTCCAAATCAGACCAAGGAACTTGAAGAAAG	2653
Qy	586	atcgtggaaatttcacataaacacattggacaaacacacagcagaaatccagattccagctc	645
Db	2654	GTCAATGGAATTCATAAATACATACAGGTCATGACTCCAGCTCAGGCTGACTTGGAAATTT	2713
Qy	646	ctagagattgccgctcagtagatgtatgaaatccggttgcacccggccaaaggacagg	705
Db	2714	CTTGAATGCCAAAAGTTGTCCATGTATGGAGTGTATCTTCACAAAAGCAAGGACTTG	2773

Qy	706	gaaggcacgaagatcaaatctgocgttgccaaacacgggaattctagtgtttcagggtttc	765
Db	2774	GAGGGAGTGGACATTATTCTCGCGCTCTGCTCCAGCGGCTTCTGCTTTACAAAGACAAG	2833
Qy	766	actaagatcaatgccttcaactgggcccagggtgcggaagctgaagcttcaagaggaagcgc	825
Db	2834	TTGAGAAATTAAACGGCTTTCTTGGCCCCAAGTGTAAAGATTCTTCCAAACCGCAGCAGC	2893
Qy	826	tttctcatcaagctccggccagatgccaatagtgcgtaccaggataaccttggaaattccctg	885
Db	2894	TTCTTCATCAAGATCCGGCTGGAGAGCAAGAAACATTATGAAAGTACCATCGGCTTCAAG	2953
Qy	886	atggccagtcgggatttctgcaagtccttcttgaaaaatctgtgttgaaacatcatgccttc	945
Db	2954	CTCCCCAGTTATCGAGCGGCCAAGAAACATATGGAAGTCTGTGTGGAGCATCACAGTTC	3013
Qy	946	ttagacttttgaagagcccaaaccaagcccaagccctctctttagccgggggtca	1005
Db	3014	TTCAGAC---TCACCTCTACAGACACCATCCCCCAAAAGCAAGTTTCTTGCCTTGGATCC	3070
Qy	1006	tcatttcggttcagtggtcggaactcagaagcag	1038
Db	3071	AAATTCGATACAGTGGCCGGACTCAAGCTCAG	3103

Search completed: December 6, 2001, 12:17:10  
Job time: 13162 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2001, 12:00:11 ; Search time 283.82 Seconds  
(without alignments)  
9469.786 Million cell updates/sec

Title: US-09-555-342A-1\_COPY\_49\_3183

Perfect score: 3135

Sequence: 1 atgggagaataagacagag.....acaaagagtctctgtgtat 3135

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3135	100.0	3442	20	Human chondrocyte-
2	3131.8	99.9	4687	21	Human cytoskeleton
3	1431.2	45.7	3094	21	Human pancreatic c
4	223.6	7.1	241	21	Human secreted exp
5	169.6	5.4	3620	21	Human cytoskeleton
6	151.6	4.8	544	22	Murine 7-transmemb
7	144	4.6	3984	18	Protein tyrosine p
8	144	4.6	3984	20	Human PTPH1 CDNA
9	143.2	4.6	2872	22	Human protein tyro
10	133.6	4.3	3166	22	Human CDNA sequenc
11	95.8	3.1	659	22	Human colon cancer

12	85	2.7	665	22	AAH07867	Human CDNA clone (
13	85	2.7	1821	22	AAH15014	Human CDNA sequenc
14	76	2.4	410	21	AAC69420	Human secreted pro
15	69.4	2.2	1156	22	AAH99833	Human protein enco
16	63.8	2.0	1626	22	AAI59140	Human polynucleoti
17	63.8	2.0	2556	22	AAI60926	Human polynucleoti
18	62	2.0	1860	22	AAF9965	Human colon carcin
19	62	2.0	5923	22	AAS02052	DNA encoding molec
20	62	2.0	5923	22	AAS02052	DNA encoding molec
21	59.2	1.9	398	21	AAH30288	Human colon cancer
22	58.4	1.9	2355	22	AAH99674	Human protein enco
23	56.8	1.8	4080	19	AAV34368	Human protein tyro
24	56.8	1.8	4080	20	AAZ11221	Human protein tyro
25	56.2	1.8	365	14	AAQ39923	Expressed Sequence
26	56.2	1.8	365	14	AAQ59335	Human brain Expres
27	54.4	1.7	1778	22	AAI60338	Human polynucleoti
28	54	1.7	550	22	AAH06244	Human CDNA clone (
29	52.8	1.7	556	22	AAH99189	Human protein enco
30	50.2	1.6	3115	21	AAC98113	Human colon cancer
31	48.2	1.5	454	22	AAI15794	Probe #5727 for ge
32	48.2	1.5	454	22	AAI37576	Probe #6262 used t
33	48.2	1.5	1730	22	AAI62819	Human CDNA SEQ ID
34	48.2	1.5	1816	22	AAI58552	Human polynucleoti
35	47.4	1.5	2400	17	AAT42859	Choline oxidase ge
36	47.2	1.5	160	22	AAI25012	Probe #14945 for g
37	47.2	1.5	160	22	AAI50750	Probe #19436 used
38	46.6	1.5	1696	22	AAH16459	Human CDNA sequenc
39	44.8	1.5	2400	18	AAT75000	Choline oxidase ge
40	44.6	1.4	347	21	AAC02360	Human secreted pro
41	44.6	1.4	625	22	AAH07308	Human CDNA clone (
42	44.4	1.4	342	21	AAF08796	Fusarium venenatum
43	44.2	1.4	10732	21	AAH10594	Gene encoding a su
44	44	1.4	463	15	AAQ68827	hgt4 clone showing
45	42	1.3	4610	21	AAA27896	Human GEF containi

#### ALIGNMENTS

RESULT 1

AAAX79183	
ID	AAAX79183 standard; DNA; 3442 BP.
XX	
AC	AAAX79183;
XX	
DT	17-AUG-1999 (first entry)
XX	
DE	Human chondrocyte-derived gene CDEP.
XX	
KW	Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;
KW	Dbl homology domain; plectstrin homology domain; rheumatoid 'arthritis;
XX	drug; ss.
XX	
OS	Homo sapiens.
XX	
PN	WO9928458-Al.
XX	
PD	10-JUN-1999.
XX	
PF	27-NOV-1998; 98WO-JP05348.
XX	
PR	27-NOV-1997; 97JP-0342060.
XX	
XX	(CHUS ) CHUGAI SEIYAKU KK.
PI	Kato Y, Kawamoto T, Koyano Y;
XX	
DR	WPI; 1999-371117/31.
XX	
DR	P-PSDB; AAY07482.
XX	
PT	Protein CDEP expressed in differentiated chondrocytes, and gene
PT	encoding it
XX	

PS Claim 5; Fig 1; 59pp; Japanese.

XX This sequence represents the coding region for a protein (CDRP) expressed  
CC in differentiated human foetal chondrocytes, which contains an ezrin-like  
CC domain, a dbl homology (DH) domain and a pleckstrin homology (PH) domain.  
CC The nucleic acid or protein can be used in the investigation and  
CC treatment of cancers and arthritic diseases (including chronic rheumatoid  
CC arthritis), or for screening of candidate anticancer drugs.

XX  
SQ Sequence 3442 BP; 864 A; 952 C; 927 G; 699 T; 0 other;

Query Match		100.0%;	Score 3135;	DB 20;	Length 3442;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 3135;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	1	atggggaaatagacagagggccgacccacaggatcacgctggggcccccggaaaattcg	60		
DB	49	atgggaaatagacagagggccgacccacaggatcacgctggggcccccggaaaattcg	108		
QY	61	gggacgtacaccttgaaacgtgacagagccgaccccaacaccttcagaaaaactcgtg	120		
DB	109	gggacgtacaccttgaaacgtgacagagccgaccccaacaccttcagaaaaactcgtg	168		
QY	121	tccatcaaaatccagatgctggatgacacccagagggcatttgaagtccacaaaagact	180		
DB	169	tccatcaaaatccagatgctggatgacacccagagggcatttgaagtccacaaaagact	228		
QY	181	cctggaaagtgctgctgatgacgtttgcaaacacctcaacctctgaaagtgactat	240		
DB	229	cctggaaagtgctgctgatgacgtttgcaaacacctcaacctctgaaagtgactat	288		
QY	241	ttggcctcgagtttctgatcacaaaaagatcacggttgctggtatctctcaaaacc	300		
DB	289	ttggcctcgagtttctgatcacaaaaagatcacggttgctggtatctctcaaaacc	348		
QY	301	attgtgaacagattagaaggccaaagcagttgtgttaagtgtgtgaaattcttt	360		
DB	349	attgtgaacagattagaaggccaaagcagttgtgttaagtgtgtgaaattcttt	408		
QY	361	ccgctgaccacacacaaactcaagaagaactcaaaagttacctgtctcgctgcagggtg	420		
DB	409	ccgctgaccacacacaaactcaagaagaactcaaaagttacctgtctcgctgcagggtg	468		
QY	421	aagcaggacttgctcaagcgagttgacgtgtaatgaccacagcgacgtctcttgatt	480		
DB	469	aagcaggacttgctcaagcgagttgacgtgtaatgaccacagcgacgtctcttgatt	528		
QY	481	tcacacattgtgcaatctgagattggggatttttgatgaagccttggacagagacactta	540		
DB	529	tcacacattgtgcaatctgagattggggatttttgatgaagccttggacagagacactta	588		
QY	541	gcaaaaaataatacatatactcagcagacgcgaactagagagacaaaatctggaatttcac	600		
DB	589	gcaaaaaataatacatatactcagcagacgcgaactagagagacaaaatctggaatttcac	648		
QY	601	cataaacacattggacaaacacagacagaatcagatttccagctctcagagattccctt	660		
DB	649	cataaacacattggacaaacacagacagaatcagatttccagctctcagagattccctt	708		
QY	661	cggtcagatgctatggaatccggttgcccccggccaaaggacaggaagcagcaagatc	720		
DB	709	cggtcagatgctatggaatccggttgcccccggccaaaggacaggaagcagcaagatc	768		
QY	721	aatctggcgttgccaaacggaaattctagttttcagggtttcactaagatcaatgcc	780		
DB	769	aatctggcgttgccaaacggaaattctagttttcagggtttcactaagatcaatgcc	828		
QY	781	ttcaactgggccaaggtgcggagctgagcttcaagagaagcgctttctcatcaagctc	840		
DB	829	ttcaactgggccaaggtgcggagctgagcttcaagagaagcgctttctcatcaagctc	888		
QY	841	cgccagatgccaatagtcgtaccaggataccttggaattctctgatggccagtcgggat	900		

DB	889	cgccagatgccaaatagtcgtaccaggataaccttggaattctctgatggccagtcggat	948		
QY	901	ttctgcaagtcctcttgaaaaatctgttgtaacatcagctctcttttagacttttgaa	960		
DB	949	ttctgcaagtcctcttgaaaaatctgttgtaacatcagctctcttttagacttttgaa	1008		
QY	961	gagcccaaaccaagcccaagccctcctcttttagcgggggtcattctcgttcagt	1020		
DB	1009	gagcccaaaccaagcccaagccctcctcttttagcgggggtcattctcgttcagt	1068		
QY	1021	ggctggactcagaagcaggttctcgactatgtttaagaaggagacataagaaggtgcag	1080		
DB	1069	ggctggactcagaagcaggttctcgactatgtttaagaaggagacataagaaggtgcag	1128		
QY	1081	tttgaagaagaacacacagaagatttctatccggagccttcttcacagctacagaa	1140		
DB	1129	tttgaagaagaacacacagaagatttctatccggagccttcttcacagctacagaa	1188		
QY	1141	ctgaattcggaaagtctggagcagctcagcagacacagccttacctttggagaaggt	1200		
DB	1189	ctgaattcggaaagtctggagcagctcagcagacacagccttacctttggagaaggt	1248		
QY	1201	gcggaattctccagggggccagagctcggcgaggaaaagaaacgaaggtttccgcggg	1260		
DB	1249	gcggaattctccagggggccagagctcggcgaggaaaagaaacgaaggtttccgcggg	1308		
QY	1261	gagccgggttcgaccccgagcctcgcgagagagaagcccgcggttaacagagcgcg	1320		
DB	1309	gagccgggttcgaccccgagcctcgcgagagagaagcccgcggttaacagagcgcg	1368		
QY	1321	gacggagccgctcgcgcccagggaggaagagagaggtcttaaggatagaccacag	1380		
DB	1369	gacggagccgctcgcgcccagggaggaagagagagaggtcttaaggatagaccacag	1428		
QY	1381	cagagttaacctcagcccccgagcccaagcagaggtcctcgactggcagtcctacctt	1440		
DB	1429	cagagttaacctcagcccccgagcccaagcagaggtcctcgactggcagtcctacctt	1488		
QY	1441	tcgagctctctgtaactcagggggagtgcccttgcacacgtgaccttgctctcc	1500		
DB	1489	tcgagctctctgtaactcagggggagtgcccttgcacacgtgaccttgctctcc	1548		
QY	1501	aacctgagcccgacacccagcaggtcctccttgatcagcccgctgctgaatgaccag	1560		
DB	1549	aacctgagcccgacacccagcaggtcctccttgatcagcccgctgctgaatgaccag	1608		
QY	1561	gctgcccccgagcgacgatgaggtgagggccggaggaagagattcccaactgataaa	1620		
DB	1609	gctgcccccgagcgacgatgaggtgagggccggaggaagagattcccaactgataaa	1668		
QY	1621	gcgtactctatagctaagaagtgtctcacccgagcgaacatctctgaaggtctcgaa	1680		
DB	1669	gcgtactctatagctaagaagtgtctcacccgagcgaacatctctgaaggtctcgaa	1728		
QY	1681	gttatacctctggttttcagagcacagtgcgcaagagagacccatgcccgaagcactg	1740		
DB	1729	gttatacctctggttttcagagcacagtgcgcaagagagacccatgcccgaagcactg	1788		
QY	1741	aaaagtctcatattcccgaaatttgaaacctttgcacaaatttcataactaatttctcaag	1800		
DB	1789	aaaagtctcatattcccgaaattttgaaacctttgcacaaatttcataactaatttctcaag	1848		
QY	1801	gaaattgacaaagcacttgcctctgtgggaagccgctcaaatgccaaatcaagattac	1860		
DB	1849	gaaattgacaaagcacttgcctctgtgggaagccgctcaaatgccaaatcaagattac	1908		
QY	1861	caagaatcggcgatgctgctgaagaacatttcagggtgagcagcctggcggtcac	1920		
DB	1909	caagaatcggcgatgctgctgaagaacatttcagggtgagcagcctggcggtcac	1968		
QY	1921	ctgtgggaagcacagcgagccttggggcccttggaatggaaatcaagagctccccggcg	1980		



Db 1969 ctgtgtgaagcacagcgagccttgaggcccttgaggatggaatgaagatcccgcg 2028  
Qy 1981 ctggagaacttctgagagactttgagctgcagaagggtgtgttaactaccgcctcaacacc 2040  
Db 2029 ctggagaacttctgagagactttgagctgcagaagggtgtgttaactaccgcctcaacacc 2088  
Qy 2041 ttccctctgcccactgacccgctcatgctactacaagcaggtctctgagcgctgtgc 2100  
Db 2089 ttccctctgcccactgacccgctcatgctactacaagcaggtctctgagcgctgtgc 2148  
Qy 2101 aaacaccaccgcccagagccacgcccacttccaggtccgagccgcttggcagagatc 2160  
Db 2149 aaacaccaccgcccagagccacgcccacttccaggtccgagccgcttggcagagatc 2208  
Qy 2161 acggagatggtggcacagctccacgtaacgatacgaatgataagatggagaattccagaagctg 2220  
Db 2209 acggagatggtggcacagctccacgtaacgatacgaatgataagatggagaattccagaagctg 2268  
Qy 2221 cacgaactcaagaagattgattgacattgacattctgtgttccgggaaggagttc 2280  
Db 2269 cacgaactcaagaagattgattgacattgacattctgtgttccgggaaggagttc 2328  
Qy 2281 atccgtctggcagcctcagcaagctctcggggaagggtccacgacgcatgtcttc 2340  
Db 2329 atccgtctggcagcctcagcaagctctcggggaagggtccacgacgcatgtcttc 2388  
Qy 2341 ctgttcaacgacgtcctgtatatacacgagccggggtcagcgccctccaatcagtttaa 2400  
Db 2389 ctgttcaacgacgtcctgtatatacacgagccggggtcagcgccctccaatcagtttaa 2448  
Qy 2401 gtccacgggacgctccgctctatgcatgacgattgagagagagcgaagcagtgggg 2460  
Db 2449 gtccacgggacgctccgctctatgcatgacgattgagagagagcgaagcagtgggg 2508  
Qy 2461 gtgcccactgctgacctccgggcccagcgcgagtcacatcatcgtgcccagttct 2520  
Db 2509 gtgcccactgctgacctccgggcccagcgcgagtcacatcatcgtgcccagttct 2568  
Qy 2521 cggctccgagatggagaagtggttgagacatccagatggccattgacctggcgagagaag 2580  
Db 2569 cggctccgagatggagaagtggttgagacatccagatggccattgacctggcgagagaag 2628  
Qy 2581 aqacacgcccgccttgagttctgtgcccagcagccccctgacaaagtcacctgat 2640  
Db 2629 aqacacgcccgccttgagttctgtgcccagcagccccctgacaaagtcacctgat 2688  
Qy 2641 gaagccacggcgtgacacagagtcagagagatgacctgagcgctcgcacatcgctg 2700  
Db 2689 gaagccacggcgtgacacagagtcagagagatgacctgagcgctcgcacatcgctg 2748  
Qy 2701 gagccagggcccgcacccgggcaacaaatggtgacgtgtgtgtgttcacaaactctgc 2760  
Db 2749 gagccagggcccgcacccgggcaacaaatggtgacgtgtgtgttcacaaactctgc 2808  
Qy 2761 agcgtctccatggtgacttcagatcagtcagtgaggagaatcagttctggaacctgtg 2820  
Db 2809 agcgtctccatggtgacttcagatcagtcagtgaggagaatcagttctggaacctgtg 2868  
Qy 2821 aggaattcaaaaaacagcaacgggtggcagaagctgtgtgtgttcacaaactctgc 2880  
Db 2869 aggaattcaaaaaacagcaacgggtggcagaagctgtgtgtgttcacaaactctgc 2928  
Qy 2881 ctgttcttcaaaatacacacaggaacataatccccctggccagcgtcctctgctcggc 2940  
Db 2929 ctgttcttcaaaatacacacaggaacataatccccctggccagcgtcctctgctcggc 2988  
Qy 2941 tactcgtccacatccctctgagtcgagacataatccagaaagacatacgtgttcaagctg 3000  
Db 2989 tactcgtccacatccctctgagtcgagacataatccagaaagacatacgtgttcaagctg 3048  
Qy 3001 cacttcaagtcctcactactactactcagggcggaagcaggtacacgttcaagaggtgg 3060  
Db 3049 cacttcaagtcctcactactactactcagggcggaagcaggtacacgttcaagaggtgg 3108

Qy 3061 atggaagtgtatccgcagtgccaccagctgctgcgacccccacgctgtgttgagccacaaa 3120  
Db 3109 atggaagtgtatccgcagtgccaccagctgctgcgacccccacgctgtgttgagccacaaa 3168  
Qy 3121 gagtctctgtgtat 3135  
Db 3169 gagtctctgtgtat 3183

## RESULT 2

AAA08582  
ID AAA08582 standard; DNA; 4687 BP.  
XX  
AC AAA08582;  
XX  
DT 19-JUL-2000 (first entry)  
XX  
DE Human cytoskeleton associated protein 2 (CYSKP-2) coding sequence.  
XX  
KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;  
KW autoimmunity; inflammatory; vesicle trafficking; neurological;  
KW cardiovascular; cell motility; reproductive; muscle disorder; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 90..3227  
FT /\*tag= a  
FT /product= CYSKP-2  
XX  
PN WO200017355-A2.  
XX  
PD 30-MAR-2000.  
XX  
PF 17-SEP-1999; 99WO-US21565.  
XX  
PR 18-SEP-1998; 98US-0172226.  
PR 27-APR-1999; 98US-0131321.  
XX  
PA (INCY-) INCYTE PHARM INC.  
XX  
PI Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;  
PI Guegler KJ, Patterson C, Azimzai Y, Baughn MR;  
XX  
DR WPI; 2000-283582/24.  
DR P-PSDB; AAY91947.  
XX  
PT Human cytoskeleton associated proteins, used to treat cell  
PT proliferative, autoimmune/inflammatory, vesicle trafficking,  
PT neurological, cell motility, reproductive and muscle disorders  
XX  
PS Claim 9; Page 101-102; 113pp; English.  
XX  
CC AAA08581-96 encode human cytoskeleton associated proteins 1 to 16  
CC (CYSKP-1 to CYSKP-16) respectively. The sequences can be used to treat  
CC and diagnose cancer and cell proliferative, autoimmune/inflammatory,  
CC vesicle trafficking, neurological, cardiovascular, cell motility,  
CC reproductive and muscle disorders. Pharmaceutical compositions  
CC containing CYSKP-2 to CYSKP-16 can be used to treat or prevent disorders  
CC associated with decreased expression or activity of CYSKP (claimed), for  
CC example, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,  
CC cancers, autoimmune/ant inflammatory disorders such as allergies, anemia,  
CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,  
CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma  
CC and trauma. CYSKP antagonists can be used to treat or prevent a  
CC disorder associated with increased expression or activity of CYSKP  
CC (claimed).  
XX  
SQ Sequence 4687 BP; 1179 A; 1261 C; 1196 G; 1051 T; 0 other;

Query Match 99.9%; Score 3131.8; DB 21; Length 4687;

Best Local Similarity 99.9%; Pred. No. 0;				
Matches 3133; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
QY	1	atggagaaatagacagagccgacccacagatcacgactgaggcccgcaaaattcg	60	
Db	90	atgggaaatagacagagccgacccacagatcacgactgaggcccgcaaaattcg	149	
QY	61	gggacagtcaccttgaaacgtggacagagccgcccacacacctcagagaaactcgtg	120	
Db	150	gggacagtcaccttgaaacgtggacagagccgcccacacacctcagagaaactcgtg	209	
QY	121	tcacataaatccagatgctgatgacaccccgaggagcatttgaagtccacaagagct	180	
Db	210	tcacataaatccagatgctgatgacaccccgaggagcatttgaagtccacaagagct	269	
QY	181	cctgggaaggtgctgctgatgacgtttgcaaccacctcaacctcgtggaagtgactat	240	
Db	270	cctgggaaggtgctgctgatgacgtttgcaaccacctcaacctcgtggaagtgactat	329	
QY	241	tttggcctcagtttctctgatcacaaaagatcaacggtggtggtggtggtggtggtggt	300	
Db	330	tttggcctcagtttctctgatcacaaaagatcaacggtggtggtggtggtggtggtggt	389	
QY	301	attgtgaacagattagaaggccaagcacgtttgttttaagtttgtggtgaaattcttt	360	
Db	390	attgtgaacagattagaaggccaagcacgtttgttttaagtttgtggtgaaattcttt	449	
QY	361	cgcctgac	420	
Db	450	cgcctgac	509	
QY	421	aagcagacttgctcaagcgaggttgacgttaatgacacacagcgacgtctcttgatt	480	
Db	510	aagcagacttgctcaagcgaggttgacgttaatgacacacagcgacgtctcttgatt	569	
QY	481	tcacacattgtgcaatctctgagattgggatttttgatgaagccttgacagagacactta	540	
Db	570	tcacacattgtgcaatctctgagattgggatttttgatgaagccttgacagagacactta	629	
QY	541	gcaaaaaataatacatcctcagcaagcagcactagagacaaaatcgtgaaattcac	600	
Db	630	gcaaaaaataatacatcctcagcaagcagcactagagacaaaatcgtgaaattcac	689	
QY	601	cataaccacattggacaaacacacagacagatcagatttccagctctcagagatgccgt	660	
Db	690	cataaccacattggacaaacacacagacagatcagatttccagctctcagagatgccgt	749	
QY	661	cggctagagatgtatggaatccggttgccacccggcccaaggacaggaagccacgaagtc	720	
Db	750	cggctagagatgtatggaatccggttgccacccggcccaaggacaggaagccacgaagtc	809	
QY	721	aacttgccgttgccaacacgggaattctagttttcagggtttcactaagatcaatgcc	780	
Db	810	aacttgccgttgccaacacgggaattctagttttcagggtttcactaagatcaatgcc	869	
QY	781	ttcaactggcccaaggtgcggaagctgagcttcaagagaaagcgtttctcatcaagctc	840	
Db	870	ttcaactggcccaaggtgcggaagctgagcttcaagagaaagcgtttctcatcaagctc	929	
QY	841	cggccagatgccaatagtgcgtaccaggaataccttggaaattcctgatggccagtcgggat	900	
Db	930	cggccagatgccaatagtgcgtaccaggaataccttggaaattcctgatggccagtcgggat	989	
QY	901	tcttgaaatcctcttggaataatcgtgtgaacatcatgctctctttagaacttttgaa	960	
Db	990	tcttgaaatcctcttggaataatcgtgtgaacatcatgctctctttagaacttttgaa	1049	
QY	961	gagcccaaaccaagcccaagccgctcctcttagccgggggtcatcatttcgggttcagt	1020	
Db	1050	gagcccaaaccaagcccaagccgctcctcttagccgggggtcatcatttcgggttcagt	1109	
QY	1021	ggtcgactcagaagcaggttctcactatgtttaagaaggaggacataaagaggtgcag	1080	

Db	1110	ggtcgactcagaagcaggttctcactactgtttaagaaggaggacataaagaggtgcag	1169	
QY	1081	tttgaagaagacacacagcaagattcattctatccgagagcttgccttcacagctacagaa	1140	
Db	1170	tttgaagaagacacacagcaagattcattctatccgagagcttgccttcacagctacagaa	1229	
QY	1141	ctgaattcgaagtcggtgagcagctctcagcagagcaccagccttacattttgagaagtc	1200	
Db	1230	ctgaattcgaagtcggtgagcagctctcagcagagcaccagccttacattttgagaagtc	1289	
QY	1201	gccgaatctccggggccagagctgcggcgaggagaaagaaacccgaagtttccgcggg	1260	
Db	1290	gccgaatctccggggccagagctgcggcgaggagaaagaaacccgaagtttccgcggg	1349	
QY	1261	gagcggggttcacccagcctcgccgagagaaagcccgcgggtaacaagcagcg	1320	
Db	1350	gagcggggttcacccagcctcgccgagagaaagcccgcgggtaacaagcagcg	1409	
QY	1321	gagcagccgctcgccgcccagaggaagagggaggtcggttaagatagagaccag	1380	
Db	1410	gagcagccgctcgccgcccagaggaagagggaggtcggttaagatagagaccag	1469	
QY	1381	cagagtaaacctcagcccccgagccaagcacaggtcctcactgactggcagtcacactt	1440	
Db	1470	cagagtaaacctcagcccccgagccaagcacaggtcctcactgactggcagtcacactt	1529	
QY	1441	tccgagctctgtgaactcgagggggagtgggcccttgccaaactgaccttgcctccc	1500	
Db	1530	tccgagctctgtgaactcgagggggagtgggcccttgccaaactgaccttgcctccc	1589	
QY	1501	aacctgagccccgacacacacagcagcctcctccttgatcagccccgctgctgaatgaccag	1560	
Db	1590	aacctgagccccgacacacacagcagcctcctccttgatcagccccgctgctgaatgaccag	1649	
QY	1561	gctgcccccgagcgacgatgagatgagggccggaggaagagattcccaactgataa	1620	
Db	1650	gctgcccccgagcgacgatgagatgagggccggaggaagagattcccaactgataa	1709	
QY	1621	gcgtactctatagctaaggaagtgtctaccacgcagcgaacatactgaagatctcgaa	1680	
Db	1710	gcgtactctatagctaaggaagtgtctaccacgcagcgaacatactgaagatctcgaa	1769	
QY	1681	gttatacctctggttttcagagcacagtgcgaaagagagcgccatgcgcgaagcactg	1740	
Db	1770	gttatacctctggttttcagagcacagtgcgaaagagagcgccatgcgcgaagcactg	1829	
QY	1741	aaaagtctcatatccggaatttgaacctttgcacaaatttcaataatttttccaag	1800	
Db	1830	aaaagtctcatatccggaatttgaacctttgcacaaatttcaataatttttccaag	1889	
QY	1801	gaaattgagcaacgacttgcctgtgggaagccgctcaaatgcccacatcagagattac	1860	
Db	1890	gaaattgagcaacgacttgcctgtgggaagccgctcaaatgcccacatcagagattac	1949	
QY	1861	caaaatacggcgatgtcatgctgaagaacatttcaggggcatgaagcacctgcggtcac	1920	
Db	1950	caaaatacggcgatgtcatgctgaagaacatttcaggggcatgaagcacctgcggtcac	2009	
QY	1921	ctgtgggaacacagcgagccttggagccctggagaaatggaaatcaagagctcccggcg	1980	
Db	2010	ctgtgggaacacagcgagccttggagccctggagaaatggaaatcaagagctcccggcg	2069	
QY	1981	ctgggaactctcagagacttgcagctgcagaaggtgtgttaacctacgcctcaacacc	2040	
Db	2070	ctgggaactctcagagacttgcagctgcagaaggtgtgttaacctacgcctcaacacc	2129	
QY	2041	ttcctcctcgccactgcacccggtcattgcactacaagcaggtccttgagcggtgtgc	2100	
Db	2130	ttcctcctcgccactgcacccggtcattgcactacaagcaggtccttgagcggtgtgc	2189	
QY	2101	aaacacaccccgccgagccaagcgcacttcagggactgcgagcgcctttggcagagatc	2160	
Db	2190	aaacacaccccgccgagccaagcgcacttcagggactgcgagcgcctttggcagagatc	2249	

Qy	2161	acggagatggtggcagctccacggtacgtatgatcaaga	tggagaatttccagaagctg	2220
Db	2250			2309
		acggagatggtggcagagctccacggtacgtatgatcaaga	tggagaatttccagaagctg	
Qy	2221	cacgaactcaagaagatttgattggcattgacaatcttggttccgggaaagggagtttc	2280	
Db	2310			2369
		cacgaactcaagaagatttgattggcattgacaatcttggttccgggaaagggagtttc		
Qy	2281	atccgtctgggcagctccagcctcctcggggaaggggtcccagcagcgcatgtttcttc	2340	
Db	2370			2429
		atccgtctgggcagcctcagcaagctctcggggaaggggtcccagcagcgcatgtttcttc		
Qy	2341	ctgttcaacgacgctctcgtctatacacagacgcgggggtgcagcgctccaatcagttttaaa	2400	
Db	2430			2489
		ctgttcaacgacgctctcgtctatacacagacgcgggggtgcagcgctccaatcagttttaaa		
Qy	2401	gtccacgggcagctcccgtctctatggcatgacgattgagagagcgaagcagcagtggggg	2460	
Db	2490			2549
		gtccacgggcagctcccgtctctatggcatgacgattgagagagcgaagcagcagtggggg		
Qy	2461	gtgcccactgocctgacctccgggcccagcggcagtcacatcatctgtggccgcagttct	2520	
Db	2550			2609
		gtgcccactgocctgacctccgggcccagcggcagtcacatcatctgtggccgcagttct		
Qy	2521	cgttcgcagatggagaagtggtgtgaggaatccagatggccattgacctggcgggagaag	2580	
Db	2610			2669
		cgttcgcagatggagaagtggtgtgaggaatccagatggccattgacctggcgggagaag		
Qy	2581	agcagcagccccctcgtgattctctggccagcagccccctgacacaagtcacctgat	2640	
Db	2670			2729
		agcagcagccccctcgtgattctctggccagcagccccctgacacaagtcacctgat		
Qy	2641	gaagccaccggcctgaccagagtcagagatgacctgagcgcctcgcgcacatcgctg	2700	
Db	2730			2789
		gaagccaccggcctgaccagagtcagagatgacctgagcgcctcgcgcacatcgctg		
Qy	2701	gagcgcagggccccgcaccgcggcaacacaaatggtcacgtgtgctggcaccgcaacac	2760	
Db	2790			2849
		gagcgcagggccccgcaccgcggcaacacaaatggtcacgtgtgctggcaccgcaacac		
Qy	2761	agcgtctccatggtggacttcagatcgcagtgagaaatcaagttgtctgaaacctgctg	2820	
Db	2850			2909
		agcgtctccatggtggacttcagatcgcagtgagaaatcaagttgtctgaaacctgctg		
Qy	2821	aggaaattcaaaaaagcaacgggtggcagagaagctgtgggtggtgtgttcaacaaactctgc	2880	
Db	2910			2969
		aggaaattcaaaaaagcaacgggtggcagagaagctgtgggtggtgtgttcaacaaactctgc		
Qy	2881	ctgttcttcaacaaatcacaccaggaacaatcatcccttgcacgctgcctctgctcgcc	2940	
Db	2970			3029
		ctgttcttcaacaaatcacaccaggaacaatcatcccttgcacgctgcctctgctcgcc		
Qy	2941	tactgctcaccaatccctctctgagtcgagaaatccagaagaagactacgttgttcaagctg	3000	
Db	3030			3089
		tactgctcaccaatccctctctgagtcgagaaatccagaagaagactacgttgttcaagctg		
Qy	3001	cactcaagtcccagcttactacttcaggggcggaagcagatcacagtttcgaaaggtgg	3060	
Db	3090			3149
		cactcaagtcccagcttactacttcaggggcggaagcagatcacagtttcgaaaggtgg		
Qy	3061	atggaagtgatccgagtgccaccagctctgctcgcgacccccacgtgtgtgagccacaaa	3120	
Db	3150			3209
		atggaagtgatccgagtgccaccagctctgctcgcgacccccacgtgtgtgagtcacaaa		
Qy	3121	gagtcctctgtgtat	3195	
Db	3210			
		gagtcctctgtgtat	3224	

### RESULTS

AAC98992

ID AAC98992 standard; cDNA; 3094 BP.

[illegible]

AAC98992:

09-MAR-2001 (first entry)

Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:220.

Human; pancreas; pancreatic cancer; pancreatic cancer antigen; detection; diagnosis; identification; histologic; cytotaxic; neuroprotective; nontropic; immunomodulatory; relaxant; contraceptive; gynecological; antiinflammatory; cardiant; gene therapy; chromosome mapping; linkage analysis; tissue identification; tissue typing; forensic; neural; immune system; muscular; reproductive; gastrointestinal; pulmonary; cardiovascular; renal; proliferative; ss.

*Homo sapiens*.

WO200055320-A1.

21-SEP-2000

08-MAR-2000: 2000WO-US05989.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA. Ruben SM:

WPI: 2000-579444/54.

F-PSDB; AAB34227.

new nucleic acid that is a pancreatic cancer antigen for preventing, treating, or ameliorating a medical condition, particular pancreatic cancer, or for use in assays for diagnosing a pathological condition -

Claim 1: Page 664-665: 1379pp: English.

AAC98773 to AAC99231 encode the human pancreatic cancer associated

AB954466. The human pancreatic cancer antigens have cytostatic, neuroprotective, neurotropic, immunomodulatory, relaxant, contraceptive, synaesthetic, cardant and antiinflammatory activities, and can be used in gene therapy. The polynucleotide and proteins can be used for preventing, treating, or ameliorating a medical condition or in assays for diagnosing a pathological condition or a susceptibility to one in a subject. Binding partners to the proteins and the activity of the proteins can be identified. The pancreatic cancer antigens can be used to detect, treat or prevent pancreatic disorders, especially cancer.

Antibodies and analogues to the antigens can be screened for. The pancreatic cancer antigen polynucleotides can be used to design nucleic acid hybridisation probes that can be used in chromosome mapping, linkage analysis, tissue identification and/or typing and a variety of forensic and diagnostic methods. The proteins can be used to generate antibodies which are used to purify, detect and target the polypeptides, including both *in vivo* and *in vitro* diagnostic and therapeutic methods. The proteins can be used to treat or prevent neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal or proliferative disorders. AAC9232 to AAC9240 and AAC85467 represent sequences used in the exemplification of the present invention.

Sequence 3094 BP: 736 A: 849 C: 755 G: 749 T: 5 other:

Query Match 45.7%; Score 1431.2; DB 21; Length 3094;  
Best Local Similarity 99.8%; Pred. NO. 0;

QY 1692 gtggtttcagagcacagtgaacaaaggagcgccatgccggagcactgaaaagtctcat 1751  
|||||  
ph 1693 ggcatttcaaacgacatgagcagaagagcgccatccccgaagcactaaaaagtctcat 121

Qv 1752 attcccqaattttgaacctttqcacaaatttcátaataattttctcaaggaaattgagca 1811

Db 122 atcccgaaattgaaaccttgcacaaatttcatactaatcttctcaaggaaattgagca 181  
QY 1812 acgacttgccctgtgggaagccgctcaaatgccaaatcagagattaccacaaagaatcgg 1871  
Db 182 acgacttgccctgtgggaagccgctcaaatgccaaatcagagattaccacaaagaatcgg 241  
QY 1872 cgtatgctatgctgaagaacattcagggtcatgaagcaccctggcggctcaccctgtggaagca 1931  
Db 242 cgtatgctatgctgaagaacattcagggtcatgaagcaccctggcggctcaccctgtggaagca 301  
QY 1932 cagcgaagcccttgagccctgagaaatgaaatcaagagctcccgccgctgagaaactt 1991  
Db 302 cagcgaagcccttgagccctgagaaatgaaatcaagagctcccgccgctgagaaactt 361  
QY 1992 ctccagagacttgagctgcagaagtggtgtacctaccgctcaaacaccttctcctcgg 2051  
Db 362 ctccagagacttgagctgcagaagtggtgtacctaccgctcaaacaccttctcctcgg 421  
QY 2052 gccactgcaccgctcatgcactacaagcaggtccctggagcgtgtgcaaacaccaccc 2111  
Db 422 gccactgcaccgctcatgcactacaagcaggtccctggagcgtgtgcaaacaccaccc 481  
QY 2112 gccagaccacgcgaacttcaggactccgagccgcttttggcagagatcacggagatggt 2171  
Db 482 gccagaccacgcgaacttcaggactccgagccgcttttggcagagatcacggagatggt 541  
QY 2172 ggcacagctccagctacgatgatcaagatggaagatttccagaagctgcacgaactcaa 2231  
Db 542 ggcacagctccagctacgatgatcaagatggaagatttccagaagctgcacgaactcaa 601  
QY 2232 gaaagatttgattggcattgacaatctgtgttcgcgggaaggaggttccatccgtctgg 2291  
Db 602 gaaagatttgattggcattgacaatctgtgttcgcgggaaggaggttccatccgtctgg 661  
QY 2292 cagcctcagcaagctctcggggaagggtcccgagcagcagatgttcttcgtttcaacga 2351  
Db 662 cagcctcagcaagctctcggggaagggtcccgagcagcagatgttcttcgtttcaacga 721  
QY 2352 cgtccctgctatacacagccggggtgcagcgcctccaatcagtttaaaagtcacagga 2411  
Db 722 cgtccctgctatacacagccggggtgcagcgcctccaatcagtttaaaagtcacagga 781  
QY 2412 gctccgctctatggatgacgattgaggagagcgaagcagatgggggtgcccaactg 2471  
Db 782 gctccgctctatggatgacgattgaggagagcgaagcagatgggggtgcccaactg 841  
QY 2472 cctgacccctcgggcccagcgcagtcacatcgttgccgcagttctcgttcccgagat 2531  
Db 842 cctgacccctcgggcccagcgcagtcacatcgttgccgcagttctcgttcccgagat 901  
QY 2532 ggagaagtggttgagacatccagatggccattgacctggcgggagaaagcagcagccc 2591  
Db 902 ggagaagtggttgagacatccagatggccattgacctggcgggagaaagcagcagccc 961  
QY 2592 cgccctgagttccttgccagcagcccccctgacacaaagtcacctgatgaagccaccgc 2651  
Db 962 cgccctgagttccttgccagcagcccccctgacacaaagtcacctgatgaagccaccgc 1021  
QY 2652 ggctgaccagagtcagagatgacatgagcgcctcgcgcacatcgtggagcgcagcc 2711  
Db 1022 ggctgaccagagtcagagatgacatgagcgcctcgcgcacatcgtggagcgcagcc 1080  
QY 2712 ccgcaccgcggcaacacaaatggtgcagtggtgctggcaccgcgaacacagcgtctccat 2771  
Db 1081 ccgcaccgcggcaacacaaatggtgcagtggtgctggcaccgcgaacacacagcgtctccat 1140  
QY 2772 ggtggacttcagcatcgcagtgagaaatcagttgtgtgaaacctgctgaggaattcaa 2831  
Db 1141 ggtggacttcagcatcgcagtgagaaatcagttgtgtgaaacctgctgaggaattcaa 1200  
QY 2832 aaacagcaacgggtgacagaagctgtgggtggtgtgttcacaaaccttctgctgttctcta 2891  
Human secreted expressed sequence tag SEQ ID NO:890.

## RESULT 4

AAA42150/c

ID AAA42150 standard; cDNA; 241 BP.

XX AAA42150;

XX 21-AUG-2000 (first entry)

Human secreted expressed sequence tag SEQ ID NO:890.

XX Human; mouse; xenopus; rat; secreted expressed sequence tag; SEST;

KW expressed sequence tag; ESR; probe; chemotactic; proliferative;

KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;

KW thrombolytic; antinflammatory; cytostatic; antibacterial; antifungal;

KW antiviral; antidiabetic; antiasthmatic; vulnery; antiparkinsonian;

KW antiulcer; osteopathic; neuroprotective; nootropic; antipsoriatic;

KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;

KW vacciner; autoimmune disorder; multiple sclerosis; allergic condition;

KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;

KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;

KW central nervous system disorder; Alzheimer's disease; stroke;

KW Parkinson's disease; Huntington's disease; coagulation disorder;

KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;

KW tumour; infection; depression; psoriasis; ss.

XX Homo sapiens.

XX OS

XX PN

XX WO200021990-A1.

XX PD

XX 20-APR-2000.

XX PF

XX 15-OCT-1999;

XX 99WO-US24205.

XX PR

XX 15-OCT-1998;

XX 98US-0104435.

XX XX

XX (GEMY ) GENETICS INST INC.

XX PA

XX XX

XX PI

XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;

XX PI

XX Merberg D, Treacy M;

XX XX

XX WPI; 2000-317937/27.

XX DR

XX Isolated polynucleotides, and encoded proteins, comprising secreted

XX PT

XX expressed sequence tags (ESTs), useful for treating various disorders

XX PT

XX such as autoimmune, infectious, and central nervous system disorders -

XX XX

XX Claim 1; Page 354; 618pp; English.

XX PS

XX AAA1261 to AAA43419 represent specifically claimed secreted expressed

XX CC

XX sequence tags (ESTs), isolated from human, mouse, xenopus and rat



Db 1206 aaaccacaggggttatgccagcagcaagctgattcccaagttcttagaaaaatgcaag 1265  
 QY 661 cggctagagatgatgaatccggttgaccgcggcccaagcagcaggaagcagcaagatc 720  
 Db 1266 agccttccatgatggtgtgacctacatcatgccaagagactcagaaggtggtgacatc 1325  
 QY 721 aatctggccgttgccaacacgggaattcttagttcaggtttcactaagatcaatgcc 780  
 Db 1326 aagctggcggtgtgctaatgacttctcatttacaagacagactgcgaatcaatcgt 1385  
 QY 781 ttaactggcccaaggtgcggaagctgagcttcaagaggaagcgtttctcacaagctc 840  
 Db 1386 ttgcttgccgaaaatcttataaaattctctataaacgcagtaacttctacattaaagtc 1445  
 QY 841 cgccagatccaatagtgctaccagataccttggaattctctgattgcccagtcggat 900  
 Db 1446 agaccggcagagctggaacggttgagagaccattggaattcgaactgcgaaccaccgg 1505  
 QY 901 tctgcaagctctctggaatactgtgttgaacatcatgctcttcttagactttttgaa 960  
 Db 1506 gcagcgaagagactggaagtgctggtgagcactacttctacaggtgtgtt--- 1562  
 QY 961 gagcccaaaccaagcccaagccgctctctttagccgggggtcactcattcgttcagt 1020  
 Db 1563 tctccagagcagcaccacaaagcgaagttctgacttggtggtccaaattctgctatagt 1622  
 QY 1021 ggtcggactcagaagcag 1038  
 Db 1623 ggcgcacccaagcag 1640

RESULT 6  
 AAH98023  
 ID AAH98023 standard; DNA; 544 BP.  
 AC AAH98023;  
 XX  
 XX  
 DT 10-OCT-2001 (first entry)  
 XX  
 DE Murine 7-transmembrane G-protein coupled receptor coding sequence #267.  
 XX  
 KW Murine; stromal stem cell; signalling; vaccine; 7TM-GPCR;  
 KW 7-transmembrane G-protein coupled protein receptor; ds.  
 XX  
 OS Mus sp.  
 XX  
 PN WO200160999-A1.  
 XX  
 PD 23-AUG-2001.  
 XX  
 PF 14-FEB-2001; 2001WO-US04700.  
 XX  
 PR 14-FEB-2000; 2000US-0182377.  
 XX  
 PA (IMCL-) IMCLONE SYSTEMS INC.  
 PA (UYPR-) UNIV PRINCETON.  
 XX  
 PI Lemischka IR, Witte L, Pereira DS;  
 XX  
 DR WPI; 2001-522596/57.  
 XX  
 PT DNA Sequences encoding 7-transmembrane G-protein coupled protein  
 PT receptors characteristic of hematopoietic stem cells, useful for  
 PT treating leukemia -  
 XX  
 PS Claim 1; Page 122; 176pp; English.  
 CC  
 CC The present invention relates to murine coding sequences for  
 CC 7-transmembrane G-protein coupled protein receptors (7TM-GPCRs). The  
 CC present sequence is one such murine 7TM-GPCR coding sequence. The present  
 CC sequence was derived from stromal stem cells. The present sequence  
 CC and its corresponding protein are useful in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate 7TM-GPCR expression.

CC 7TM-GPCRs identify specific signalling molecules, to activate an  
 CC effector-signalling cascade that triggers an intracellular response and  
 CC eventually a biological effect.  
 XX  
 SQ Sequence 544 BP; 143 A; 131 C; 150 G; 120 T; 0 other;

Query Match 4.8%; Score 151.6; DB 22; Length 544;  
 Best Local Similarity 56.3%; Pred. No. 1.3e-30;  
 Matches 304; Conservative 0; Mismatches 234; Indels 2; Gaps 1;

QY 2231 agaaagatttgattgacattgtgacaaattctgtgttccgggaagaggttccatcgtctg 2290  
 Db 4 agcgagacctggtcggtgtagaaacctcattgtctctgggaggttccatcgtcgagg 63  
 QY 2291 gaagctcagcaagctctcgggaaggggtccagcagcagcatgttcttctcttcaacg 2350  
 Db 64 gctgctgcacaagctcacaagaaggcgtgcacaagagaggttttcttctctcag 123  
 QY 2351 acgtctctgtatcacagcgcgggggtgagcgcctcccaatcagtttaaagtcacagggc 2410  
 Db 124 atattgtgtgtatcaaacgaaagtgtcacagagccagtcatttccggatcgtggt 183  
 QY 2411 agctcccgctctatgacatgacattgagagagcagagagtggtgggtggtccctcact 2470  
 Db 184 tcttccactcgtggtggtgtagaagaagtgagaatgaatggtgtcttctcctcatt 243  
 QY 2471 gctgaccttcggggccagcgcagtcctcactcagtcggtggtccagttctcgtccgaga 2530  
 Db 244 gcttcacctctatgcagctcagaacaaattgtgtgagcagcagcactcgtgtagaa 303  
 QY 2531 tgagaagtggttgaggacatccagatggcattgacctggtggaagagcagcagcc 2590  
 Db 304 agaaaagtgtgatgcaggacctgaatgcagcaatccaagcagcagcactcgtgact 363  
 QY 2591 cgcctcctgagttcctcggcagcagccccctgcacaaagtccctctgtatgaagccacog 2650  
 Db 364 caccctcagtgctggtggagggccgtgtatactgtaccctctagatctctgtatgaag 423  
 QY 2651 cggctacccagagtcagagtgacctgagcgcctcgcgcacatcgtgagcgcagg 2710  
 Db 424 tctctctggaagatcagaagatggtcagaggaacccgggtccctcgtggaggggaac--ag 481  
 QY 2711 cccgcacccggcgaacacaatgtgtcacgtgtgtgtggtggaacccacacacagcgtctcca 2770  
 Db 482 ccagcaccggggcaatacaacaatgcattgtgtgtgtgtgtgtgtgtgtgtgtgtgtcca 541

RESULT 7  
 AAT58627  
 ID AAT58627 standard; cDNA; 3984 BP.  
 XX  
 AC AAT58627;  
 XX  
 DT 22-MAY-1997 (first entry)  
 XX  
 DE Protein tyrosine phosphatase cDNA.  
 XX  
 KW Protein tyrosine phosphatase; PTPH1; focal adhesion;  
 KW protein tyrosine kinase; malignancy; cancer; gene therapy;  
 KW retrovirus; vector; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 24..2765  
 FT /\*tag= a  
 XX  
 XX US5595911-A.  
 XX  
 XX 21-JAN-1997.  
 PD  
 XX 14-MAR-1990; 90US-0494036.  
 PF





localise to focal adhesions and is therefore potentially useful in the treatment of cancer. Overexpression of pPrl can be used to counter the effects of oncogenic protein tyrosine kinases such as those of transforming viruses and for interfering with or reversing cell transformation. This would provide a means of preventing or reversing abnormally high levels of phosphotyrosine associated with any disease or condition such as preventing or reversing malignancy associated with the activity of a protein tyrosine kinase.

Sequence 3984 BP; 1086 A; 954 C; 930 G; 1014 T; 0 other;

Query Match 4.6%; Score 144; DB 20; Length 3984;  
Best Local Similarity 50.8%; Pred. NO. 4e-28;  
Matches 423; Conservative 0; Mismatches 400; Indels 9; Gaps 3;

QY	126	caaaatccagatgctggtgatgacacccaggagcatttggaaattccacaaagagctcctgg	185
Db	111		
Db	116	cagatccacttttagtggcgtgtacagacctttaaagtactaacaagacactgg	175
QY	186	gaagtgctgctgagtcagcttgaaccacctcaacctcgtgaagtgactatttgg	245
Db	111		
Db	176	ccagttcttctgatatggtgcacacccacctgggtgtagctgaaaggaaatatcttgg	235
QY	246	cctcgagtcttctgatcacaaaaagatacacgggtg---tggtcgtgatctcctaaacccat	302
Db	111		
Db	236	tttacgcatgatgacgactccgtggactccttagatgctggaagcaagcaaacccat	295
QY	303	tgtgaacacagattagaaggccaaagcagcttggttgtaagtttgttgtaaatcttccc	362
Db	111		
Db	296	caggaaagcagtaaaaggaggttccccctgtacacctgaatttcagtaagatctttat	355
QY	363	gcctgcacacacaaactccaaagaagaactcacaaagttacactgttcgcgtgcaggtgaa	422
Db	111		
Db	356	acctgattcccaacacactgagcaagaacaaacaggcaactgtattcttacaactgaa	415
QY	423	gcaggacttggtctaaaggcaggttgacggttaatgacacccagcgcagctctctgattc	482
Db	111		
Db	416	gatggattttgcgaaggaaaggttaacctgcctctaaactcagcagtggtctagcgtc	475
QY	483	acacattgtcaactctgagattgggtgattttgatgaagcctt--ggacagagagcaact	539
Db	111		
Db	476	ctatgcgcgtcaaatctcatatttggagactaaattcttcatacatcatccaggctatct	535
QY	540	agcaaaaaataaatcacatacctcagcaagacgcactagaggacaaaaatcgtggaattcca	599
Db	111		
Db	536	ttccgatagtcactttatacccgatacaaaatgaggactttttaacaaaatcgaaatctct	595
QY	600	ccat---aacacatttggacaaacacacagcagaatactagatttccagctccttagagattgc	656
Db	111		
Db	596	gcatgagcacacagtggtctaaacaaatcagaagcagaatccgtgctatatcaacatagc	655
QY	657	ccgtcggtgtagagatgatggaaatccggttgaccccgccaaagcacaggggaagcacgaa	716
Db	111		
Db	656	gcggaccctgcactctatggagtagaactgacagtggttaggtatctgcacaaatttaga	715
QY	717	gatacattcgtcggttgccaacacggaattctctagtgttcagggtttccactaaagatcaa	776
Db	111		
Db	716	ccataatgatlgaaattgcttccgcgggtgtgtcgtgtacccgaaatacatattgcacaa	775
QY	777	tgccttcaactgggccaaggtgcggaagctgagotccaagagggaagcgcttctcatcaa	836
Db	111		
Db	776	tttcta tcttgggtgaaacttccaaattcttccaaaggaaaaagttcttcataca	835
QY	837	gtctccggccagatgccaatagtcgtaccaggataccttggaattctctgatggccagtcg	896
Db	111		
Db	836	tcagcgacagaaaacaggctgaatccagggaacatatgtggccttcaacatgctgaaatta	895
QY	897	ggatttctgcaagtctcttctggaaaaatctgtgttgaacatacatcctctcttt	948
Db	111		
Db	896	ccgatcttgaaaaacttgtgaaatcctgtgttgagaccacatacgttctcttt	947

RESULTS

RESOLUT 3  
AAS08887  
ID AAS08887 standard: cDNA: 2872 BP.

AA  
AC  
AAS08887;

XX DT 26-SEP-2001 (first entry)

Human protein tyrosine phosphatase (PTPase) DNA #2.

Protein tyrosine phosphatase; pTPase; human; nervous system; dementia; Alzheimer's disease; Huntington's disease; Parkinson's disease; autism; Tourette's syndrome; obsessive compulsive disorder; schizophrenia; mania; cardiovascular disorder; atherosclerosis; myocarditis; Addison's disease; autoimmune disorder; rheumatoid arthritis; multiple sclerosis; vitiligo; ankylosing spondylitis; Sjogren's syndrome; asthma; atopic dermatitis; diabetes mellitus; inflammatory disorder; pancreatitis; sarcoidosis; ss; allogeneic transplant rejection; multiple myeloma; Burkitt's lymphoma; leukaemia; cancer; neoplasm.

XX Homo sapiens. OS

Key	Location/Qualifiers
FH	230..1339
FT	CDS

```

ET      /tag=a
ET      /product="Human PTPase #2"

```

AX WO200153530-A1.  
PN

PD 26-JUL-2001.

AA  
PF 17-JAN-2001; 2001WO-US01563.AA  
PR 18-JAN-2000; 2000US-0176306.

PA (HUMA-) HUMAN GENOME SCI INC.

Shi Y, Ruben SM:

AA  
DR  
WPI; 2001-442265/47.

XX  
DR P-PSDB; AAU04486.

New human protein tyrosine phosphatase polypeptide useful for treatment of disorders including disorder of the nervous system -

PS Claim 1; Page 282-283; 290pp; English.

The sequence represents DNA encoding a human protein tyrosine phosphatase (PTPase) polypeptide. PTPase proteins and nucleic acids are useful for preventing, treating or ameliorating a medical condition. The medical conditions are disorders of the central and peripheral nervous system (e.g. Alzheimer's disease, Huntington's disease, Parkinson's disease, Tourette's syndrome, obsessive compulsive disorder, schizophrenia, mania, dementia and autism), cardiovascular disorders (e.g. atherosclerosis and myocarditis), autoimmune disorders (e.g. rheumatoid arthritis, multiple sclerosis, ankylosing spondylitis, Addison's disease, Sjogren's syndrome, vitiligo, asthma, atopic dermatitis and diabetes mellitus), inflammatory disorders (e.g. pancreatitis, sarcoidosis and allogeneic transplant rejection) and cancers and neoplasms (e.g. leukaemia, Burkitt's lymphoma and multiple myeloma). The presence or absence of a mutation in the nucleic acid encoding the protein allows for diagnosis of or susceptibility to a pathological condition. The sequences are also useful in screening for agonists and antagonists of PTPase activity.

Sequence 2872 BP: 840 A: 575 C: 601 G: 856 T: 0 other;  
XX

Query Match	Score	DB	Length
Best Local Similarity	4.6%	143.2	22
Matches	50.3%	Pred. No. 5.5e-28	
Conservative	0	Mismatches 468	
Indels	16	Gaps	5

QY 83 gacagaagccgccccaacaccttcaggaaaactcgtgtcccatcaaaatccagatgctgg 142



[illegible]

RESULT 10

AAH17787

ID AAH17787 standard; cDNA; 3166 BP.

XX

AC AAH17787;

XX

[illegible]

26-JUN-2001 (first entry)  
Human cDNA sequence SEQ ID NO:17434.  
Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
Homo sapiens.  
EP1074617-A2.  
07-FEB-2001.  
28-JUL-2000; 2000EP-0116126.  
29-JUL-1999; 99JP-0248036.  
27-AUG-1999; 99JP-0300253.  
11-JAN-2000; 2000JP-0118776.  
02-MAY-2000; 2000JP-0183767.  
09-JUN-2000; 2000JP-0241899.  
(HELI-) HELIX RES INST.  
Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
WPI: 2001-318749/34.

primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8; SEO ID 17434; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesising polynucleotides particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH02166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 3166 BP; 922 A; 671 C; 699 G; 874 T; 0 other;

Query Match 4.3%; Score 133.6; DB 22; Length 3166;  
Best Local Similarity 49.5%; Pred. NO. 2.2e-25;  
Matches 495; Conservative 0; Mismatches 484; Indels 21; Gaps 5;

QY 137 tgctgatatgacaccagaggcatttgaagtccacaaagagctctctgggaaggtgctgc 196

Db 288 ttctggatgttactgatgttagtgtggacttgccaaaaaacgcaaggacaaagtgtg 347

QY 197 tggatgcagtttgcacaaccctcgtggaagtgactattttgggcctcgagttc 256

Db 348 ttgatcagattatgtaccacctggacctgattgaaagcgactattttgggtctgagattta 407

QY 257 ctgatcacaacaaagatcacggtgtggctggatctcctaaacccattgtgaaacagatta 316

Db	408	tggaattcagcacaaagtagcacattggttggaTggtacaaaagcatcaaaaagcaagtaa	467
Qy	317	gaaggccaagacagctgtgtttaaagtttgttgtaaatctttccgcgtgaccacacac	376
Db	468	aaatgtgtcaacctattgtctgactcttcaggttaagtgtttattcttcagaacaaata	527
Qy	377	aactccaagaagaactcaaaaggtacctgttcgcgtgcaggtgaagcaggaacttggctc	436
Db	528	acctctggaggagtaaacccggtatttattgtcttcagTtaaaacaagatatattcca	587
Qy	437	aaggcaggttgacgtgtaatatgacacacagcgcagctcttgaatttcacacatttgcagt	496
Db	588	gtggaataattagactgtccotttgaTcacagcgtgcaatttggcagcttataattctgcaag	647
Qy	497	ctgagattggggattttgatgaagcctttggacag--agagcaactttagcaaaaataaat	553
Db	648	ctgaacttggtagctatgattctgtgacatagtagctctgaactgtctcagaagttcagat	707
Qy	554	acataccc--tcagcaagacgactagaggaacaaaactcgtggaaTtccaccataaccaca	610
Db	708	tcgtgcattatcagctgaagagatggaaactggctatttttgagaaTtgggaaggaataca	767
Qy	611	ttggacaaacacagcagaaatcagatttccagctcctagagatgTccogtcogctagaga	670
Db	768	gaagtcaaacacagcacaggtctgaaccaaTtatctgaataaaggccaaatTggctagaaa	827
Qy	671	tgtatgaaTccgggttgcccccggccaaggaacaggaaggcacgaagatcaatctggccg	730
Db	828	tgtagtgggttgatagcatgtgtgTccaagctagagatTggaaTgactatagtttgggac	887
Qy	731	ttgccaaacacgggaaTcttagtgtttccagggttttcaactaaTgatacaatgcttcaactggg	790
Db	888	taacaccaacaggagtcctgtttttggaaggagataccaaaatTgcttattttttTggc	947
Qy	791	ccaaggtgcggaagctgagcttcaagaggaagcgtttctcatcaagctccggccagatg	850
Db	948	cgaagataaacagatTcgatttttaagaaTaaaattaaccttggTgtttagaagatg	1007
Qy	851	ccaat-----agtgtaccaggataccttggaaTtccgtgatggccagtcggattctt	904
Db	1008	atgatcagggcgaagaacaggacaatacatattgtcttttagactgataTccaagaagcat	1067
Qy	905	gcaagctctcttgaaaaatctgtgttgaacataTgccttcttttagacttttgaagagc	964
Db	1068	gcaaacatttatTgaaatgtgtgtggagcatcatgtctttctccgctctcgagcccg	1127
Qy	965	ccaaaccacaagcccaagc--ccgtcctcttttagccggggtcatctatttcggttcagt	1021
Db	1128	tccaaaagagTtctcatcgatcagatttatTcgactaggatcacagatttagataTatagt	1187
Qy	1022	gtcggactcagaagcaggttctcgactatTttaagaaggaggagacataagaaggtgcagt	1081
Db	1188	ggaaaacaggtatcagaccacaaaaacaaTaaagcaagaagatcaa-----catcct	1241
Qy	1082	ttgaagggaagcagcagatcatcttatctccggagcct	1121
Db	1242	ttgaagaagTcccaagcaacagataTcttagacggaactct	1281

RESULT	11
AAH33845	
ID	AAH33845 standard; cDNA; 659 BP.
XX	
XX	AAH33845;
XX	AC
XX	03-SEP-2001 (first entry)
DT	
DE	Human colon cancer antigen encoding cDNA SEQ ID NO:901.

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
KW colorectal carcinoma; ss.  
KW XX

OS	Homo sapiens.	
XX	WO2001122920-A2.	
PN	05-APR-2001.	
XX	28-SEP-2000; 2000WO-US26524.	
XX	29-SEP-1999; 99US-0157137.	
PR	03-NOV-1999; 99US-0163280.	
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA	Ruben SM, Barash SC, Birse CE, Rosen CA;	
XX	WPI: 2001-235357/24.	
DR	P-PSDB; AAG74414.	
XX	Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers -	
PT	Claim 1; Page 2820; 9803pp; English.	
PS	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytostatic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAB77789 represent sequences used in the exemplification of the present invention.	
CC	N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922.	
CC	Sequence 659 BP; 170 A; 147 C; 172 G; 163 T; 7 other;	
XX		

Query Match 3.1%; Score 95.8; DB 22; Length 659;  
Best Local Similarity 53.6%; Pred. No. 1.3e-15;  
Matches 199; Conservative 0; Mismatches 172; Indels 0; Gaps 0;

QY	137	tgtctgattgacacccaggagggcatttgaagtctccaaagagagctcttggaaggtgtctgc	196
DG	58	tgtctgcagcggccgaagtggagcgtggacctgcgaacaatgcacaaaggccagatttgt	117
QY	197	tggatgcagtttgcacaaccactcaacctcgttggaaggtgactattttggcctcgagtttc	256
DG	118	tgtatcagatttgtaccacttggaccttggaaacagatttactttggcctccagttcc	177
QY	257	ctgatcacaaaaagatcacaggtgtggctggatctctctaaaaccatttgtaaacagatta	316
DG	178	tgcacttgcacaggttgcgcactggcgtgatactgccaaaccctaaaaaagcatga	237
QY	317	gaagccaaaacagctgtgtgttaaagtttgggtgaattcttcgcctgcaccacacac	376
DG	238	aatctggacctcttatgctttacactttcgagttgaaatctactattcttcagaaccacaa	297
QY	377	aactccaagaagaactcacaagttacctgttcgcgtgcaggtgaagcaggactttggctc	436
DG	298	accttcgtgaggagtttacaaggtacctgtttgtttacaactcaggcatgatctcttt	357
QY	437	aaggcaggttgacgtgtaatatgacacacagcgcagctctcttgatttacacatttgtcaat	496
DG	358	cttggaaaaattgaaatgcctcttatgaaacagctgtggaaattagctgtctctgtctacaag	417





CC increase or decrease storage capabilities. AAC69390 to AAC69398 and  
CC AAB38118 represent sequences used in the exemplification of the present  
CC invention.  
XX  
SQ Sequence 410 BP; 100 A; 93 C; 105 G; 112 T; 0 other;

Query Match 2.4%; Score 76; DB 21; Length 410;  
Best Local Similarity 94.0%; Pred. No. 2.1e-10;  
Matches 79; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1013 ggttcaggtgctgacacagacaggtctcactatgttaagaaggagcataaga 1072

Db 53 gtttcaggtgctgacacagacaggtctcactatgttaagaaggagcataaga 112

QY 1073 aggtgcagttgaaaggaagcaca 1096

Db 113 aggtgcagttgaaaggaagcaca 136

## RESULT 15

AAH99833

ID AAH99833 standard; cDNA; 1156 BP.

XX

AC AAH99833;

XX

DT 16-OCT-2001 (first entry)

XX

DE Human protein encoding cDNA sequence SEQ ID NO:568.

XX

KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;  
KW antiinflammatory; antirheumatic; antiarthritic; immunosuppressive;  
KW antibacterial; endocrine; cardiant; central nervous system; virucide;  
KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;  
KW antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;  
KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;  
KW neuroprotective; antidepressant; nootropic; antiparkinsonian; infection;  
KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;  
KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;  
KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;  
KW genetic disease; haematopoietic disorder; platelet disorder; asthma;  
KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;  
KW allergic rhinitis; diabetes; multiple sclerosis; depression;  
KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;  
KW neurological disorder;; ss.

XX Homo sapiens.

OS

XX WO200153455-A2.

PN

XX 26-JUL-2001.

PD

XX 22-DEC-2000; 2000WO-US35017.

XX

XX 23-DEC-1999; 99US-0471275.

PR

XX 21-JAN-2000; 2000US-0488725.

PR

XX 25-APR-2000; 2000US-0552317.

XX

XX (HYSE-) HYSEQ INC.

PA

XX Tang YT, Liu C, Drmanac RT;

PI

XX WPI; 2001-457603/49.

DR

XX P-PSDB; AAM25892.

XX

XX Isolated human polynucleotides encoding polypeptides, useful for the

PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX

XX Claim 1; Page 674; 1217pp; English.

XX

XX AAH99166 to AAH99904 encode the human proteins given in AAM25225 to

CC AAM25963. The proteins can have activities based on the tissues and

CC cells they are expressed in, such as: antiinflammatory; antirheumatic;

CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant;  
CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;  
CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;  
CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;  
CC antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic;  
CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides  
CC encoding them can be used in gene therapy, antisense therapy and vaccine  
CC production. The proteins and polynucleotides are useful for screening for  
CC agonists or antagonists of a protein and for the treatment and diagnosis  
CC of disorders associated with the activity of a protein e.g. inflammation,  
CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,  
CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal  
CC infections, autoimmunity, genetic diseases, haematopoietic disorders,  
CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,  
CC osteoporosis, severe combined immunodeficiency, eczema, allergic  
CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,  
CC Alzheimer's disease, Parkinson's disease, neurodegenerative and  
CC neurological disorders.  
XX  
SQ Sequence 1156 BP; 381 A; 202 C; 255 G; 318 T; 0 other;

Query Match 2.2%; Score 69.4; DB 22; Length 1156;

Best Local Similarity 50.1%; Pred. No. 2.1e-08;

Matches 262; Conservative 0; Mismatches 246; Indels 15; Gaps 3;

QY 608 acattggacaacacacagcagaaatcagattccagctcctagagattgcccgctgctag 667

Db 381 acagaggtcaacacacagcagaggtgaacaaattatctgaataaagcaaatgctag 440

QY 668 agatgtatgaatccggtgtgcacccggcgaagacagagggacacgaagatacaatctgg 727

Db 441 aaatgtatgggttgatgatcgtgtgtcgaagctagatgggaatgactatagtttg 500

QY 728 ccgttgccaacacgggaattcttagtttcagggtttcactaagatcaatgcctcaact 787

Db 501 gactaacacacacagaggtcctgttttgaaggagataccaaaattggctatttttt 560

QY 788 gggcccaaggtgcggaagctgagcttcaagaggaagcgtcttctcatcaagctccggccag 847

Db 561 ggcgaagataacacagattggttttaagaagaataataaaccttgggttgtagaag 620

QY 848 atgccaat-----agtcgtaccagataaccttgggaattcctctgtagccagtcgggatt 901

Db 621 atgatgatcaggggcaagaaacaggaacatacatttctttagactggatccatccaaag 680

QY 902 tctgcaagtccttctgaaaaatctgtgtgaacatacgtccttcttagactttttgaag 961

Db 581 catgcaaacatttatggaaatgtgtgtggagcaatcgtttcttccgcttcgagggcc 740

QY 962 agcccaaaccaagcccaagc---ccgtcctctttagccggggtcatcatttcgttca 1018

Db 741 ccgtccaaaagagttctcatcagcagatttattcagctaggatcacgatttagata 800

QY 1019 gtggtcggactcagaagcaggttctcagctatgtttaagaaggagagcataaagaagtgc 1078

Db 801 gtgggaaaacagagatcatcagaccacaaaacccaataagaagaagatcaaa-----cat 854

QY 1079 agtttaagaaggacacgaagattcttctatcccgagacct 1121

Db 855 ccttgaagaaggcccgcaagcaacgatatcttagacgaacct 897

Search completed: December 6, 2001, 12:01:00

Job time: 11372 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2001, 14:32:44 ; Search time 102.73 Seconds  
(without alignments)  
6911.404 Million cell updates/sec

Title: US-09-555-342A-1\_COPY\_49\_3183

Perfect score: 3135

Sequence: 1 atgggagaatagacagag.....acaaagagctctgtgtat 3135

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PTCUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	56.8	1.8	4080	2	US-08-446-345-35
2	53.8	1.7	7218	1	US-08-232-463-14
3	47.4	1.5	2400	4	US-08-930-001-1
4	44.4	1.4	463	1	US-07-752-101A-67
5	42.2	1.3	441529	4	US-09-103-840A-1
6	41.2	1.3	4403765	4	US-09-103-840A-2
7	40.8	1.3	1598	4	US-08-211-682-24
8	40.8	1.3	4776	2	US-08-852-401-1
9	40.4	1.3	4791	4	US-08-949-155-49
10	40	1.3	2277	1	US-08-676-967-5
11	40	1.3	2277	1	US-08-676-974-5
12	40	1.3	2277	2	US-09-098-487-5
13	38.6	1.2	2574	2	US-08-677-734A-8
14	38.4	1.2	4267	4	US-08-949-155-51
15	38.2	1.2	1053	6	5352575-6
16	37.6	1.2	2619	4	US-08-467-822-19
17	37.6	1.2	2619	4	US-08-432-697-19
18	37.6	1.2	2619	4	US-08-466-248-19
19	37.6	1.2	2634	1	US-08-196-218-31
20	37.6	1.2	2634	1	US-08-681-953-31
21	37.6	1.2	3468	1	US-07-951-715A-2
22	37.6	1.2	3468	1	US-07-951-715A-4
23	37.6	1.2	3468	2	US-08-459-448A-2
24	37.6	1.2	3468	2	US-08-459-448A-4
25	37.6	1.2	3468	3	US-08-459-595A-2
26	37.6	1.2	3468	3	US-08-459-595A-4
27	37.6	1.2	3468	3	US-08-459-504B-2

28	37.6	1.2	3468	3	US-08-459-504B-4	Sequence 4, Appl
29	37.6	1.2	3468	3	US-08-459-444-2	Sequence 2, Appl
30	37.6	1.2	3468	3	US-08-459-444-4	Sequence 4, Appl
31	37.6	1.2	3468	3	US-09-053-549-3	Sequence 3, Appl
32	37.6	1.2	3468	3	US-09-053-549-5	Sequence 5, Appl
33	37.6	1.2	3475	4	US-09-657-481A-10	Sequence 10, Appl
34	37.6	1.2	3476	4	US-08-630-916A-47	Sequence 47, Appl
35	37.2	1.2	2745	1	US-08-363-255-1	Sequence 1, Appl
36	37.2	1.2	2745	1	US-08-363-255-13	Sequence 13, Appl
37	37.2	1.2	44377	2	US-08-804-227C-7	Sequence 7, Appl
38	37.2	1.2	44377	2	US-08-804-198-1	Sequence 1, Appl
39	36.8	1.2	2376	1	US-07-912-952-3	Sequence 3, Appl
40	36.6	1.2	1273	4	US-09-319-892-3	Sequence 3, Appl
41	36.6	1.2	1288	1	US-08-440-856A-9	Sequence 9, Appl
c 42	36.4	1.2	2824	2	US-09-010-928B-3	Sequence 3, Appl
c 43	36.4	1.2	3509	3	US-08-322-962-1	Sequence 1, Appl
c 44	36.4	1.2	3509	3	US-08-450-653-1	Sequence 1, Appl
c 45	36	1.1	2214	3	US-08-864-038A-1	Sequence 1, Appl

#### ALIGNMENTS

RESULT 1  
US-08-446-345-35  
; Sequence 35, Application US/08446345  
; Patent No. 5831009  
; GENERAL INFORMATION:  
; APPLICANT: Ullrich, Axel  
; APPLICANT: Moller, Niels P.H.  
; APPLICANT: Moller, Karin B.  
; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE  
; NUMBER OF INVENTION: PHOSPHATASES PTP-DI  
; NUMBER OF SEQUENCES: 41  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: N.Y.  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/446,345  
; FILING DATE: 22-MAY-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/234,440  
; FILING DATE: 28-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30742  
; REFERENCE/DOCKET NUMBER: 7683-054  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212)790-9090  
; TELEFAX: (212) 869-8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 35:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4080 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: both  
; TOPOLOGY: unknown  
; MOLECULE TYPE: CDNA  
US-08-446-345-35

Query Match 1.8%; Score 56.8; DB 2; Length 4080;  
Best Local Similarity 46.6%; Pred. No. 2.5e-05;









MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/852,401

CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mueller, Lisa L.  
REGISTRATION NUMBER: 38,978  
REFERENCE/DOCKET NUMBER: FER2159PO0300US  
TELEPHONE: 312-616-5400  
TELEFAX: 312-616-5460

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4776 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-852-401-1

Query Match 1.3%; Score 40.8; DB 2; Length 4776;  
Best Local Similarity 43.9%; Pred. No. 0.52;  
Matches 174; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 1964 tcaagagctcccgccggtggagactctgcagagactttgagctgcagaaggtgtgtt 2023

Db 797 TCAACCGCGCGGGGCGATCAAGAGGAGACCCGCCAGCGCGTGTGGCCGTCTCAACG 856

QY 2024 acctaccgtcaacacattctctctgcccactgcacggctcatgctactacaagcagg 2083

Db 857 AGATCGCGCACACCGCGCGCATAGGTGCGTGGCGCGCACCGGGTAAATCGGCCTCG 916

QY 2084 tctcgagcggtgtgcaaacaccaccgccagccagccagctcagaggaactgcggag 2143

Db 917 TGGTGGCGGAGGTGTCACACCGCGGTGTTCCCGCCTTCGCCAGGCGCCGTGGAGCGCGG 976

QY 2144 ccgctttgcagagatcacggatggtggcacagctccacggtacgtagatcaagatgg 2203

Db 977 CGGTGGCGCGGGCTACGGCTCTGCTGTGCAACCCCGCTCGGGATGAGCGAGGAGG 1036

QY 2204 agaattccagaagctgcacgaactcaagaagatttgattggcattgacaattctgtgg 2263

Db 1037 ACTAGTCCGGATGCTCATCGCGCGCGGTGGAGGCGATGCTTCGTGTCGCGGACA 1096

QY 2264 ttccgggaaggaggttcctctctg99cagcctcagcaagctctcggggaagggtccc 2323

Db 1097 TCGCCAACACCGAGGGCGAGCGCGATCAGCCGAGCTACTACGAGAAGCTGCTGGCG 1156

QY 2324 agcagcgcatgtctctctctgttcaacgacgtctgc 2359

Db 1157 ACGCGTGGCGATGGTCTTCGTAACGGCGCGCGC 1192

## RESULT 9

US-08-949-155-49  
Sequence 49, Application US/08949155  
Patent No. 6271436

GENERAL INFORMATION:  
APPLICANT: Piedrahita, Jorge A.  
APPLICANT: Bazer, Fuller W.  
TITLE OF INVENTION: Compositions and Methods for the  
Generation of Transgenic Animal Species  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE AND DURKEE  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX

COUNTRY: US  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/949,155  
CLASSIFICATION: 800  
PRIOR APPLICATION NUMBER: US 60/027,338  
FILING DATE: 11-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/046,094  
FILING DATE: 09-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Hibler, David W.  
REGISTRATION NUMBER: 41,071  
REFERENCE/DOCKET NUMBER: TAMK:177  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4791 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-949-155-49

Query Match 1.3%; Score 40.4; DB 4; Length 4791;  
Best Local Similarity 59.6%; Pred. No. 0.66;  
Matches 68; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 2521 cggtcogagatggagaagtgggttgaggacatccagatggccattgacctgcgcggaag 2580

Db 2866 CGCCAGGCTCTGGAGAAGAGAGAGCTGGAGCGCTGAGAGGGGAACGCTCCCTGCAGGACCAG 2925

QY 2581 agcacagcccgccctcgtgcttcctggcagcagcccccctgacacaaagLcc 2634

Db 2926 AGTCGAGCGCTCTCCCTAAGCTGCTGCCCTGCCCCCCCCCGCCACCC 2979

## RESULT 10

US-08-676-967-5  
Sequence 5, Application US/08676967  
Patent No. 5747317

GENERAL INFORMATION:  
APPLICANT: COLLINS, KATHLEEN  
TITLE OF INVENTION: Human Telomerase  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Science & Technology Law Group  
STREET: 268 Bush Street, Suite 3200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard A  
REGISTRATION NUMBER: 36,627



```

QY 2505 cgtggcgcagttcttcggttcgcagatggagaagtgggttagggacatccagatggccat 2564
      ++++++ || || || || || || || || || || || || || || || || || ||
Db 804 GAACGGCGCGTGAAGCGCCCGCCCGCCCGCCGCAAGAGCAGCAGCAGCAGGAGACAG 863
      ++++++ || || || || || || || || || || || || || || || || || ||
QY 2565 tgacctggcggagaagagcagcagcagcccgccctcgagttctggccagcagcccccctga 2624
      || || || || || || || || || || || || || || || || || || || || ||
Db 864 CGACCTGGAGGAGAGCGACACATCGACGCGCGGAGAGCTGGCCCGCAGAGCGACACCAG 923
      || || || || || || || || || || || || || || || || || || || || ||
QY 2625 caaacatccctgatgaagccacgcg 2652
      || || || || || || || || || || || || || || || || || || || || ||
Db 924 CACCGAGGAGCAGGAGGACACGCGCGTG 951
      || || || || || || || || || || || || || || || || || || || || ||

RESULT 13
US-08-677-734A-8
; Sequence 8, Application US/08677734A
; Patent No. 5871919
; GENERAL INFORMATION:
; APPLICANT: Brant, Steven R.
; APPLICANT: Yun, Chris C.H.
; APPLICANT: Donowitz, Mark
; APPLICANT: Tse, Chung-Ming
; TITLE OF INVENTION: Cloning, Tissue Distribution, and
; TITLE OF INVENTION: Functional Analysis Of The Human Na+/H+ Exchanger Isoform,
; TITLE OF INVENTION: NHE3.
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,734A
; FILING DATE: 10-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 05387.0043-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2574 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-677-734A-8

Query Match 1.2%; Score 38.6; DB 2; Length 2574;
Best Local Similarity 44.1%; Pred. No. 1.5;
Matches 161; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

QY 2272 agggagttcatccgtctgggcagctcagcaagctctcggggaaggggctccagcagcgc 2331
      || || || || || || || || || || || || || || || || || || || || ||
Db 1762 AGAGAAAATGTCACGCGTGTCTGCTGGACATGTCAGTCTCTGGAGCAGCGACGGGGAGC 1821
      || || || || || || || || || || || || || || || || || || || || ||
QY 2332 atgttctctgttccacgacgtctcgtctatacacagagccgggctgaagcctccaat 2391
      || || || || || || || || || || || || || || || || || || || || ||
Db 1822 ATCCGGGACGGGAGGACATGCTGTCAGCCACACACGCTACAGCATCTGTACAAAGCGG 1881
      || || || || || || || || || || || || || || || || || || || || ||
QY 2392 cagtttaaagtcacggcagctccgctctatgcatgaagattgaagagagcccaagac 2451
      || || || || || || || || || || || || || || || || || || || || ||

```

Db	1882	CGCAGGAGTACAAAGCATCTGTACAGCCGACACGAGCTCACGCCACCGGAGGACGAGAA	1941
Qy	2452	gagtggggggtgcccccaactgacctcccggggcccagcgagtcagtcacatctgtgccc	2511
Db	1942	CAGGACCGGGAATCTTCCACAGGACCATGCGGAAGCGCTGGAGTCCTTCAAGTCGACC	2001
Qy	2512	gccagttctcgttcgagatggagaagtgggttgaggacatccagatggccattgacctg	2571
Db	2002	AAGCTGGGGCTCAACACGAACAAGAAGGAGCGCAAGCTGTATCAAGCGGAGCGTGCCCG	2061
Qy	2572	gcggagaagcagcagcagcccccctgagttcttctggccagcagccccctgacacaag	2631
Db	2062	AAGCGAGAAACAGCAGCATCCCAATGGGAAGCTGCCCATGGAGAGCCCTGCCAGAA	2121
Qy	2632	tcgcc 2636	
Db	2122	TTCAC 2126	
RESULT 14			
US-08-949-155-51			
; Sequence 51, Application US/08949155			
; Patent No. 6271436			
; GENERAL INFORMATION:			
; APPLICANT: Piedrahita, Jorge A			
; APPLICANT: Bazer, Fuller W			
; TITLE OF INVENTION: Compositions and Methods for the			
; TITLE OF INVENTION: Generation of Transgenic Animal Species			
; NUMBER OF SEQUENCES: 51			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: ARNOLD, WHITE AND DURKEE			
; STREET: P.O. Box 4433			
; CITY: Houston			
; STATE: TX			
; COUNTRY: US			
; ZIP: 77210-4433			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: Floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: PatentIn Release #1.0, Version #1.30			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: US/08/949,155			
; FILING DATE: Concurrently Herewith			
; CLASSIFICATION: 800			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 60/027,338			
; FILING DATE: 11-OCT-1996			
; PRIOR APPLICATION DATA:			
; APPLICATION NUMBER: US 60/046,094			
; FILING DATE: 09-MAY-1997			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Hibler, David W.			
; REGISTRATION NUMBER: 41,071			
; REFERENCE/DOCKET NUMBER: TAMK:177			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: (512) 418-3000			
; TELEFAX: (713) 789-2679			
; INFORMATION FOR SEQ ID NO: 51:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 4267 base pairs			
; TYPE: nucleic acid			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
US-08-949-155-51			

Query Match 1.28; Score 38.4; DB 4; Length 4267;  
Best Local Similarity 47.28; Pred. No. 2.2;  
Matches 117; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

Qv 1327 qccgcctcgcgccaccgaggaaggaggaqggtcgttaaqgatagaccacgcagaqt 1386

```
Db 3020 GCCTCTCTGCGCCAGGAGGCTGATAGAGGAGCATGAAGGAGGTGAAGGCGCTACCGC 3079
QY 1387 aaactcaagcccccagcaagcaagcagagctccctactggcagtgctccactttccgag 1446
Db 3080 GAGGAGCTGAGGCGCGAGCTGGGCCCGCTGACCCAGGAGACGCGCGCGCTGTCCAAG 3139
QY 1447 ctgtctgtgaactcgaggggggagtgggccctgccaagtgaccttgtctcccaacctg 1506
Db 3140 GAGCTGAGGCGCGCGAGCGCCCGTGGCGCGGACATGAGGAGACGTGCGCAACCGCTTG 3199
QY 1507 agccccagaccagcagagcctctccttgatcagcccgctgtctgaatgaccaggcctgc 1566
Db 3200 GTGCTCTAGCGCAGCGAGGTGCAACAACATGTTGGCCAGACCAACCGAGGAGCTGCGGAGC 3259
QY 1567 ccccgagac 1574
Db 3260 CGCCTGGC 3267
```

## RESULT 15

5352575-6

; Patent No. 5352575

; APPLICANT: PETROVSKIS, ERIK A.; POST, LEONARD E.; TIMMINS, JAMES G.

; TITLE OF INVENTION: PSEUDORABIES VIRUS PROTEIN

; NUMBER OF SEQUENCES: 12

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/513,282

; FILING DATE: 20-APR-1990

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 100,817

; FILING DATE: 29-JUN-1987

; APPLICATION NUMBER: 886,260

; FILING DATE: 16-JUL-1986

; APPLICATION NUMBER: 784,787

; FILING DATE: 04-OCT-1985

; APPLICATION NUMBER: 801,799

; FILING DATE: 26-NOV-1985

; APPLICATION NUMBER: 844,113

; FILING DATE: 26-MAR-1986

; SEQ ID NO: 6:

; LENGTH: 1053

; 5352575-6

## Query Match

1.2%; Score 38.2; DB 6; Length 1053;

Best Local Similarity 54.0%; Pred. No. 1.2;

Matches 101; Conservative 0; Mismatches 83; Indels 3; Gaps 1;

```
QY 1215 gggccagagctgcgcgcgaggaacccgaaggtttccgcgcgggagccgggtcgca 1274
Db 564 gtggcagcagcctgcgcgtcgtgaccccgccgagagccggtgttaccacgcgcgc 623
QY 1275 cccgagc---cctgcgcagagagagcccccgcgggtgataaagcagcgagcgccgc 1331
Db 624 cccgatacagcagagccgcgcgcaccccccgcgcgcgcgcgcgcgcgcgcgcgcgc 683
QY 1332 ctgcgcgcagcagagagagagagctcgttaagatagagacccagcagagtaaac 1391
Db 684 cagccccgcctccgacgaagagagagagagagaggggggggacgacgagcgatgacccc 743
QY 1392 tcagccc 1398
Db 744 ggtgccc 750
```

Search completed: December 6, 2001, 16:58:20  
Job time: 30017 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2001, 10:34:44 ; Search time 2912.37 Seconds  
(without alignments)  
11567.217 Million cell updates/sec

Title: US-09-555-342A-1\_COPY\_49\_3183

Perfect score: 3135

Sequence: 1 atgggagaatagacagag.....acaaagagtctctgttat 3135

Scoring table:

IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_est1:\*  
11: gb\_est2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	784.8	25.0	1122	10	BE746268
2	761.2	24.3	804	11	BG747792
3	757.2	24.2	865	10	BE260677
4	755.8	24.1	784	11	BG763918
5	753	24.0	930	10	AL566821
6	745	23.8	861	11	BG764061
7	726.2	23.2	901	11	BG475554
8	716	22.8	736	11	BE793662
9	705.4	22.5	898	10	BE745887
10	704.6	22.5	835	11	BG829192
11	700.2	22.3	902	11	BE910036
12	687.6	22.9	721	10	AU132546

13	679.6	21.7	942	11	BG323704
14	678.2	21.6	891	11	BG767698
15	673	21.5	883	11	BF206296
16	660.6	21.1	981	11	BG171514
17	660	21.1	680	11	BG825738
18	628	20.0	859	10	AL537488
19	625.4	19.9	741	11	BG122769
20	616.8	19.7	690	11	BG698789
21	608.4	19.4	968	11	BF206873
22	607.6	19.4	934	11	BF314265
23	606	19.3	621	11	BG770181
24	605.6	19.3	692	11	BG116225
25	599.4	19.1	890	11	BE907778
26	598.8	19.1	704	11	BG769615
27	590.4	18.8	932	11	BF686586
28	572.8	18.3	788	11	BG750463
29	570.6	18.2	861	11	BG420356
30	551.8	17.6	954	11	BF304259
31	551	17.6	560	10	AW411445
32	550.4	17.6	582	10	AI907882
33	529.8	16.9	627	10	AV729533
34	507	16.2	507	10	AI683520
35	506	16.1	757	10	BE282962
36	491.6	15.7	994	11	BT081073
37	489.6	15.6	592	11	BG078481
38	486	15.5	719	11	BG248068
39	478.4	15.3	939	11	BF316537
40	475.4	15.2	489	10	AL121548
41	463.2	14.8	612	10	AW129611
42	458.6	14.6	596	10	AI360995
43	456.4	14.6	1002	11	BG323892
44	452.6	14.4	918	11	BF178599
45	446.8	14.3	450	10	AI690075

#### ALIGNMENTS

RESULT 1

LOCUS BE746268

DEFINITION 601579756F1 NIH\_MGC\_9 Homo sapiens cDNA clone IMAGE:3928572 5', mRNA

ACCESSION BE746268

VERSION BE746268.1

KEYWORDS EST. GI:10160260

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1122)

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-remail.nih.gov

Tissue Procurement: DCTD/DTF

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov

Plate: LLCM759 row: p column: 13

High quality sequence stop: 780.

Location/Qualifiers

1..1122

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:3928572"

/clone\_lib="NIH\_MGC\_9"

/tissue\_type="adenocarcinoma cell line"

/lab\_host="DH10B (phage-resistant)"

/note="Organ: ovary; Vector: pOTB7; Site\_1: XhoI; Site\_2:





Qy	679	atccggttgaccccgcccaaggagacagggaagcagcaagatcaatctggccgttgccaac	738
Db	242	ATCCGTTTGACCCGCGCAAGGACAGGAGCGAAGATCAATCTGGCCGTGCGCAAC	301
Qy	739	acgggaattcctagtgtttcagggtttcactaagatcaatgccttcaactgggccaaggtg	798
Db	302	ACGGGAATTCATGTGTTTCAGGGTTTCACTAAGATCAATGCTTCAACTGGGCGCAAGGTG	361
Qy	799	cggaaagctgagcttcaagaggaagcgctttctcatcaagctccggccagatgcctaagtg	858
Db	362	CGGAAGCTGAGCTTAAGAGGAGGCGCTTTCATCAAGCTCCGGCCAGATGCCAATAGT	421
Qy	859	gcgtaccaggaataccttggaaattcctgatggccagtcgggaattcttgcaagtcctctcg	918
Db	422	CGGTACCAGGATACCTTGGAAATCTGTATGCGCAGTCGGGATTTCTCAAGTCCTTCTTG	481
Qy	919	aaaatcgtgtgaacatacgtccttctttagaacttttgaagagcccaaaacaaagccc	978
Db	482	AAAATCTGTGTGAACATCATGCCCTTCTTAGACTTTTGAAGAGCCCAACCAAGSCCC	541
Qy	979	aagccgcctcttttagccgggggtcatcatttcggttcagtgctcggaactcagaagcag	1038
Db	542	AAGCCCGCTCTTTAGCGGGGGTTCATCATTTTGGTTCAGTGGTCGGACTCAGAGCAG	601
Qy	1039	gttctcgactatgtt--aaagaagaggaacataagaaggtagctttgaaggaagacaca	1096
Db	602	GTTCTCGACTATGTTTACACAGGAAGGAGACATAAGAAGTGCAGTTTGAAGGAAGCAC	661
Qy	1097	gcaagattcaattctatccggagccttgcttcacagcctacaagaactgaatt--cggaagtg	1155
Db	662	G--AAGATTCATTCTATCCGAGCCCTTGCTTCACAGCCTACAGAACTCAAAATCCGGAAGTG	720
Qy	1156	ctggagcagctctcagcagacaccagccttacatttgagaaggtgcggaatctccaggg	1215
Db	721	CTGGACGACTCTCAGCAGACACACAGCTTACATTTTGGAGAGAGTGCCGGAATCTCCAGG	780
Qy	1216	ggccagagctgcggcgagggaag	1239
Db	781	GGACAGAGCTGCCGCGAGGAAAG	804

```

RESULT 3
BE260677      865 bp      mRNA      26-OCT-2000
LOCUS        601154188F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3510381 5',
DEFINITION   mRNA sequence.
ACCESSION    BE260677
VERSION      BE260677.1 GI:9132166
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 865)
NIH-MGC http://mgc.ncl.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgabbs-r@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: LLCM194 row: o column: 22
High quality sequence stop: 730.
Location/Qualifiers
1. .865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3510381"
FEATURES
source

```

```

/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming, Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GCCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
207 a 236 c 255 g 167 t

```

Query Match	24.28;	Score	757.2;	DB	10;	Length	865;
Best Local Similarity	97.3%;	Pred.	No. 1.8e-175;				
Matches	823;	Conservative	0;	Mismatches	18;	Indels	5;
Gaps							
5;							
Qy	671	tgtatgaatccggttgcacccgcgaagacagggagacgaagaatcaatctggccg	730				
Db	1	TGTATGGNAATCCGGTGTGACCCGGCCAAAGGACAGCAAGGACCAAGATCAATCTGGCCG	60				
Qy	731	tgcacacacgggaattctagtgttccagggtttcactaaatcaatgccttcaactggg	790				
Db	61	TTGCCACACACGGGAATCTAGTGTTCAGGGTTTCACTAAGATCAATGCCTTCAACTGG	120				
Qy	791	ccaagtgcggaagctgagcttcaagaggaagcgcttctcatcaagctcgggccaagt	850				
Db	121	CCAAGTGCAGAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGCTCCGGCCAGATG	180				
Qy	851	ccaatagtcgtaccaggataccttgaattccctgatgcccagtcgggattcttgcagt	910				
Db	181	CCAATAGTGCATACAGGATACCTTGGAAATTCCTGATGCCAGTCGGGATTTCTGCAAGT	240				
Qy	911	cttctggaaatctgtgtgaacatcatgcctctctttagactctttgaaagagcccaaac	970				
Db	241	CCTTCTGGAAAACTGTGTTGAACATCATGCTTCTTTTAGACTTTTGAAGAGCCCAAC	300				
Qy	971	caagcccaagccgctctctttagccgggggttcatactcttcggttcaagtggtcggaetc	1030				
Db	301	CAAAAGCCCAAGCCGCTCCTCTTTAGCCGGGGTTCATCATTTTCGGTTCAGTGGCGAGTCT	360				
Qy	1031	agaagcaggttctcgactatgttaaagaagagagacataaagaagtcgagtttgaagaaga	1090				
Db	361	AGAGCAGGTTTCTGCATATGTTTAAAGAGGAGGACATAAAGAGTGCAGTTTGNAAAGA	420				
Qy	1091	agcacagcaagattcattctatccggagcccttgcctcacagcctcacgaactgaattcgg	1150				
Db	421	AGCACAGCAAGATTCTTCTATCCGGAGCCTTGCTTTCACGCCTACACAACTGAATTCGG	480				
Qy	1151	aagtgtgagcagtcctcagcagagcacccgcttatacttggagaagtgccgaatctc	1210				
Db	481	AAGTGTGGAGCAGTCTCAGCAGAGCACACAGCCTTACATTTTGGAGAAGGTGCCGAATCTC	540				
Qy	1211	caggggccaagctgcgcgcgagaaaggaacgaaggtttccgcgcggggagccggggt	1270				
Db	541	CAGGGGGCCAGAGCTGCCGGCAGAAAGGAACCGAAGGTTTCCGCCGGGAGCCGGGGT	600				
Qy	1271	cgcacccgagccctgcgcgcgaggaagaccgcgcgggttaacaagcagcgcgacagagccg	1330				
Db	601	CGCACCCGAGCCCTGCGCCGAGGAGAAACCCCGGGGTAAACAAGCAGCGCAGCGAGCCG	660				
Qy	1331	cctcggcgcccaaggaagagagagggctgtaagaagataggaccacagcagagtaaac	1390				
Db	661	CCTCGGCCCCACGGAGGAAGAGGAGGAGTGC-TAAGGATAGGA-CCAGCAGAGTTAAC	718				
Qy	1391	ctcagcccccagcccaagcagagctccctgact-ggcagtcctcacttccagagctg	1449				
Db	719	CTCAGCCCCCGAG-CAAGACACAGCTCCCTTGACTGGGCAGTCTCTACCTTTCGAGCTG	777				
Qy	1450	tctgtgaactcgaggggggagtgggccctgcaccaagtcgacctgtctcccaacctgagc	1509				

```
Db 778 TCTGTGACTCGCAGGGGAGTTGGCCCTCGGCCACGTAACC-TGTTTCCCAACCTGGGC 836
QY 1510 ccgac 1515
|| ||
Db 837 CCGAAC 842

RESULT 4
BG763918 784 bp mRNA EST 15-MAY-2001
LOCUS 602736888F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:486244 5',
DEFINITION mRNA sequence.
ACCESSION BG763918
VERSION BG763918.1 GI:14074571
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 784)
TITLE NIH-MGC http://mgi.nci.nih.gov/.
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-re@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1722 row: o column: 21
High quality sequence stop: 761.
FEATURES
source
location
1..784
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:486244"
/clone_lib="NIH_MGC_49"
/tissue_type="melanotic melanoma, high MDR (cell line)"
/note="Organ: skin; Vector: pORF7; Site1: XhoI; Site2:
ECORI; cDNA made by oligo-dT priming. Directionally cloned
into ECORI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library. |"
BASE COUNT 203 a 223 c 223 g 135 t
ORIGIN
Query Match 24.1%; Score 755.8; DB 11; Length 784;
Best Local Similarity 99.5%; Pred. No. 3.9e-175;
Matches 779; Conservative 0; Mismatches 2; Indels 2; Gaps 2;
QY 1123 gcttcacagcctacagaactgaattcggaagtgcgtggagcagctctcagcagagcaccagc 1182
Db 2 GCTTCACGCCTACAGAACTGAATTCGGAAGTCTGGAGAGCTCTCAGCAGACACACAGC 61
QY 1183 cttacattggagaaggtccgaatccaggggccagagctgcgcgcgaggaagaa 1242
Db 62 CTTACATTTGGAGAGGTGCCGATCTCCAGGGGGCCAGAGCTCCCGCAGGAAGAA 121
QY 1243 ccgaaggtttccgcggggagccgggtgcacccagagccctgcgcgcgaggaagccccc 1302
Db 122 CCGAAGGTTTCCCGCGGGAGCGGGGTGCGACCCCGAGCCCTGCGCGCGAGGAAGCCCC 181
QY 1303 gcgggttaacagcaggcgagcgagccgctcgccacagggaggaagaggaggtc 1362
|||||
```

```
Db 182 GCGGGTAACAAGCAGCGGAGCGCGCTCGGCCCCACGGAGGAAGAGGAGGTC 241
QY 1363 gtaaggatagaccagcagagtaaacctcagcccccagcccccagcagcagcagcagcctcctg 1422
||| |||
Db 242 GTTAAGGATAGGACCCAGCAGAGTAACCTCAGCCCCCGCCAGCCACGACAGGCTCCCTG 301
QY 1423 actggcagtcctcacccttcccgagctgtgtgaaactcaggggggagtgggccctgccc 1482
||| |||
Db 302 ACTGGCAGTCCCTACCTTTCCGAGCTGTGTGTAACCTCGCAGGGGGAGTGGCCCCCTGCC 361
QY 1483 aacgtgacctgtctctccaaacctgagccccgcagacacaaagcagcagcctctccttgatcagc 1542
||| |||
Db 362 AACGTGACCTTGTCTCCCAACCTGAGCCCCGACACCAAGCAGGCGCTCTCCCTTGATCAGC 421
QY 1543 ccgctgctgaatgaccagcctgcctgccccgcagcgacgatgagggcgaggaagaa 1602
Db 422 CCGCTGCTGAATGACGAGCGCTGCCCGGACGAGATGAGGATGAGGCGCGGAGGAAG 481
QY 1603 agattcccaactgataaagcgtactctcatagctaaagagtgctctaccaccgcagcaac 1662
||| |||
Db 482 AGATTCCCAACTGATAAAGCGTACTTTCATAGCTTAAGGAAGTGTCTACCACCGAGCGAACA 541
QY 1663 tatcgaagagatctcgaagtattcacttgcgtgtttcagagcagcagtgacaaagagac 1722
||| |||
Db 542 TATCTGAAGGATCTCGAAGTTTATCACTTCGTGGTTTCAGAGCACAGTGAGCAAAAGAGGAC 601
QY 1723 gccatgccggaagcactgaaagtctcatattcccaattttgaaactttgacacaaattt 1782
||| |||
Db 602 GCCATGCCGGAAGCACTGAAAGTCTCATATTTCCCGAATTTGAACCTTTGCACAAATTT 661
QY 1783 catactaattttctcaagaaa-ttgagcaacgacttgcctgtggaagggcgctcaca 1841
||| |||
Db 662 CATACTAAATTTCTCAAGGAAACTTCAGCAACGACTTGCCCTGTGGGAAGCGCGCTCAA 721
QY 1842 tgcccaaatcagagattaccagaatcgcgatgcatgc-tgaagaacattcagggca 1900
||| |||
Db 722 TGCCCAATCAGAGATTACCAAGAAATCGCGGATGTCATGCTTGACGAACATTCAGGCCA 781
QY 1901 tga 1903
|||
Db 782 TGA 784

RESULT 5
AL566821/c
LOCUS AL566821 LTI_FL013_Fbrnl Homo sapiens cDNA clone CS0DF025YJ01 3
DEFINITION prime, mRNA sequence.
ACCESSION AL566821
VERSION AL566821.1 GI:12919564
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 930)
AUTHORS Li.W.B., Gruber.C., Jessee.J. and Polayes.D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
location/Qualifiers
1..930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DF025YJ01"
/clone_lib="LTI_FL013_Fbrnl"
/dev_stage="pooled tissue from post conception fetuses (20
week_24 week and 26 week)"
/lab_host="DH108"
/note="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand
```

cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 186 a 249 c 283 g 211 t 1 others  
ORIGIN

Query Match 24.0%; Score 753; DB 10; Length 930;  
Best Local Similarity 97.9%; Pred. No. 2e-174;  
Matches 806; Conservative 0; Mismatches 10; Indels 7; Gaps 4;

QY 2318 gqctccagcagcagtgcttcttcctgttcaacagcgtcctgtatcacagcgggggc 2377  
|||||  
DB 930 GGCTCCAGCAGCGCATGTTCTTCCTGTTCAACGACGTCTGTATACAGCGGGGGC 871  
|||||

QY 2378 tgacggctcccaatcagtttaagtcacgggcagctcccgctctatgcatgacgattg 2437  
|||||  
DB 870 TCACGGCTCCCAATCAGTTAAAGTCCAGGCGCAGCTCCCGCTCTATGCGATGACGATTG 811  
|||||

QY 2438 agagagcgaagacagagtg9gggggtgccccactgctgacctccggggccagcgagtg 2497  
|||||  
DB 810 AGGAGAGCGAAGACAGAGT-GGGGGTGCCTGCTGAACTCCGGGGCCAGCGCGAGT 752  
|||||

QY 2498 ccatcatgtgcccagcttcgtccagatgagagtgagtgagagatccaga 2557  
|||||  
DB 751 CCATCATGTCGGCCGCCAGTTCGTGGTCGAGATGAGAAAGTGGGTGAGGACATCCAGA 692  
|||||

QY 2558 tggccattgacctggcgagaagacagcagcccgccctgagttctctggccagcagcc 2617  
|||||  
DB 691 TGGCCATTGACCTGGCGAGAGACAGCAGCCCCGCCCTGAGTCTCTGTCGCCAGCGCC 632  
|||||

QY 2618 cccctgacaaagtcctctgatgaagccacgcgctgaccagagtgatgacc 2677  
|||||  
DB 631 CCCCTGACAAAGTCCCTGTATGAAGCACCCCGCGCTGACCAGGAGTCAGAGGATGACC 572  
|||||

QY 2678 tgagcgtctcgacatgctggagccagccagcccgccagcagcggaacaaatggctc 2737  
|||||  
DB 571 TGAGGCGCTTCGGCGACATCGCTGGAGCGCGCAGG-CCCCCACC CGCGCACAAATGGTGC 513  
|||||

QY 2738 acgtgtgctggcaccgcaacacagcgtctccatggtggaacttca---gcacgcagtg 2794  
|||||  
DB 512 ACSTGTGCTGGCACCACACACAGCGTCTCCATGGTGGACTTCACGCMCCCGCCAGTGG 453  
|||||

QY 2795 agaatacgtgtcttggaacacctgctgaggaataatcaaaacagcaacggtggcagaagc 2854  
|||||  
DB 452 AGAATCAGTGTGCTGGAACCTGCTGAGGAAATTCAAAACACAGCAACGGGTGGCAGAAGC 393  
|||||

QY 2855 tdtg--gggtggttcaaaacttgcctgttcttctcaaatcacacagagacaatca 2912  
|||||  
DB 392 TGTGGCGTCGGTGTACAAATCTTGCCTGTCTTCTTCAAAATCACACAGGACAAATCA 333  
|||||

QY 2913 tcccttgcagcgtcctctgctggctactcgtccactccctctgagtcaggagaa 2972  
|||||  
DB 332 TCCCTTGCACGCTGCCTCTGCTCGGTACTCGCTACCATCCCTCTGTGTCGAGAA 273  
|||||

QY 2973 catccagaagactacgtgttcaagctgcaacttcaagtcacagtcctactacttcagggc 3032  
|||||  
DB 272 CATCCAGAAAGACTAGGTGTTCAAGTGCACCTTCAAGTCCCACTCTACTACTCTCAGGC 213  
|||||

QY 3033 gaaaagcagtagcacgttccgaagtgatgaaatgatccagtgccacagctctc 3092  
|||||  
DB 212 GGAACCGAGTAGACGTTCCGAAGGTGGATGGAAGTGATCCCGAGTGCACACAGCTCTGC 153  
|||||

QY 3093 ctccgacccccagctgttgagccacaaagagctctctgtgtat 3135  
|||||  
DB 152 CTCGGACCCACAGTGTGAGTACAAAGAGTCTCTTGTGTAT 110  
|||||

## RESULT 6

BG764061

## LOCUS

BG764061 861 bp mRNA EST 15-MAY-2001  
602737071f1 NIH\_MGC\_49 Homo sapiens cDNA clone IMAGE:4862417 5',  
DEFINITION mRNA sequence.

## ACCESSION

BG764061

## VERSION

BG764061.1 GI:14074714

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

1 (bases 1 to 861)

## AUTHORS

NIH-MGC http://mgc.nci.nih.gov/.

## TITLE

National Institutes of Health, Mammalian Gene Collection (MGC)

## JOURNAL

Unpublished (1999)

## COMMENT

Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue Procurement: ATCC/DCFD/PTP

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

http://image.llnl.gov

Plate: LCM1722 row: n column: 18

High quality sequence stop: 817.

## FEATURES

source

1..861  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4862417"  
/clone\_lib="NIH\_MGC\_49"  
/tissue\_type="melanotic melanoma, high MDR (cell line)"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: skin; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally cloned  
into EcoRI/XhoI sites using the following 5' adaptor:  
GGCAGGAG(G). Size-selected >500bp for average insert size  
1.8kb. Library constructed by Ling Hong in the laboratory  
of Gerald M. Rubin (University of California, Berkeley)  
using ZAP-cDNA synthesis kit (Stratagene) and Superscript  
II RT (Life Technologies). Note: this is a NIH\_MGC  
Library. |"

BASE COUNT 198 a 245 c 270 g 147 t 1 others

## ORIGIN

Query Match 23.8%; Score 745; DB 11; Length 861;

Best Local Similarity 96.9%; Pred. No. 1.8e-172;

Matches 791; Conservative 0; Mismatches 21; Indels 4; Gaps 3;

QY 1862 aaagaatcggcctatgctgaagaacattcaggcgcagcctgagcctccgcgctcacc 1921

|||||

DB 2 AAAGAATCGGCATGTCTATGCTGAAGAATTCAGGGCATGAAGCACCTCGCGGCTCACC 61

|||||

QY 1922 tgtgaagcacagcagcgtctggagcctggagaatggaatcaagagctcccgcgccg 1981

|||||

DB 62 TGTGGAAGCACAGCGAGCGCTTGGAGGCCCTTGAGAAATGAATCAAGAGCTCCCGCGCG 121

|||||

QY 1982 tggagaactctgcagagagactttgagctcagaagtgctgttacctacccgtccaacacct 2041

|||||

DB 122 TGGAGAACTTCTGCAGAGACTTTTGAGCTGCAGAAAGGTGTGTACCTTACCGCTCAACACCT 181

|||||

QY 2042 tctcctcgccacctgacccgctcatcactacaagcagctctggagcgtgtgca 2101

|||||

DB 182 TCTCTCTGGGGCACTGTCACCGGCTCATGCCTACAAAGAGGTCCTTGGAGCGGCTGTGCA 241

|||||

QY 2102 aacaccaccgcgcagcaccgcgcagcttcaggagactgcgcagcgcgtttggcagagatca 2161

|||||

DB 242 AACACCACCGCGGAGCCACCGCGACTTCAGGGACTTGCAGCGCGCTNTGGCAGAGATCA 301

|||||

```
QY 2162 cggagatgtggtgacagctccacggtacgatgatcgaatggagaatttccagaagtgc 2221
Db 302 CGGAGATGTGTGGCACAGCTCCACGGTACGATGATCAAGATGGAGATTTCCAGAGCTGC 361
QY 2222 acgaactcaagaagatttgatggcattgacaatttgggttccgggaaggagattca 2281
Db 362 ACGAATCTCAAGAAGATTTGATTGGCATTGACAATCTTGTTGTCGGGAAGGAGTTCA 421
QY 2282 tcgctgtggcagcctccagcaagctctcggggaagggtccagcagcgcatgttcttc 2341
Db 422 TCGCTGTGGCAGCCTCAGCAAGCTCTCGGGGAAGGGGCTCCAGCAGCGCATGTTCTTC 481
QY 2342 tgttaacgacgtctctgtatatacacgagcggggggtgacggctcccaatcagtttaaag 2401
Db 482 TGTTCACGACGCTCTGCTATACACGAGCGGGGGCTGACGGCTCCTCAATCAGTTTAAAG 541
QY 2402 tcacggggaagctcccgctctatggcattgacgattgagagagcgaagcgaagtggggg 2461
Db 542 TCCACGGGAGCTCCCGCTCTATGGCATGACGATTGAGGAGCGAAGACGAGTGGGGGG 601
QY 2462 tgcccacgtgctgacccctccggggcagcggcgagtcctcatcgtgcccagttctc 2521
Db 602 TGCCCCACTGCTGACCTTCCGGGGCAGCGGCGAGTCCATCATCTGTGGCGCCAGTCTC 661
QY 2522 ggtccgagatggagaagtgggttgagacatccagatggccattgacctggcggaag- 2580
Db 662 GPTCCGAGATGGAGAAGTGGTTGAGACATCCAGATGCCATTGACCTGGCGGAGAA 721
QY 2581 agcagcagcccgccctgagttcttgccagcagcagcccccctgacaaagtccctgat 2640
Db 722 AGCCAGCAGCCCGCCCTGAGTTCCTTGGCCAGCAGCCCCCTGACAAAGT-CCCTGAT 780
QY 2641 gaagcaccgcggtgaccagagtcagaggtgac 2676
Db 781 GAAGC--ACCGCGTGACCAAGAGTCAGAGGATGAC 814

RESULT 7
BG475554
LOCUS 502491544F1 NIH_MGC_20 Homo sapiens cDNA clone IMAGE:4620346 5',
DEFINITION mRNA sequence.
ACCESSION BG475554
VERSION BG475554.1 GI:13407833
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 901)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: ATCC/DCTD/DTP
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM1381 row: h column: 11
High quality sequence stop: 818.
Location/Qualifiers
1..901
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4620346"
/clone_lib="NIH_MGC_20"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: skin; Vector: pORF7; Site_1: XhoI; Site_2:
```

```
BASE COUNT 232 a 243 c 253 g 173 t
ORIGIN
Query Match 23.2%; Score 726.2; DB 11; Length 901;
Best Local Similarity 91.3%; Pred. No. 7.8e-168;
Matches 817; Conservative 0; Mismatches 68; Indels 10; Gaps 4;
QY 1532 ccttgatcagccgcgtgctgaatgaccagcctgccccggagcgaagcagatgagatgaag 1591
Db 2 CTTTGATCAGCCGCTGCTGAATGACCAAGCTTGCCTCCCGGACGAGATGAGGATGAGG 61
QY 1592 gccggaggaagagattcccaactgataaagcgtacttcatagctaaaggaagtgtctacca 1651
Db 62 GCGGAGGAAGAGATTCCCACTGATAAAGCGTACTTTCATAGCTAAGGAAGTGTCTACCA 121
QY 1652 ccgagcgaacatctatgaaggatctcgaagtattcacttcgtggtttcagagcacagtga 1711
Db 122 CCGAGCGAATATCTGAAGGATCTCGAAGTATATCACTTCGTGTTTCAGAGCACAGTGA 181
QY 1712 gcaagagagacccatgcccgaagcactaaaaagtctcatattcccgaaatttgaacctt 1771
Db 182 GCAAGAGAGACGCCATGCGGAGACTGAAAAGTCTCATATTTCCCGAATTTTGAACCTT 241
QY 1772 tgcaaaaatttatactaattttcgaagaaattgagcaagcaactgcccgtggggaag 1831
Db 242 TGCACAAATTTCTACTAATTTCTCAAGAAATTTGAGCAACGACTTGCCTGTGGGAAG 301
QY 1832 gccgctcaaatgccaaatcagagattaccaaagaatcgcgatgtcatctgctgaagaaca 1891
Db 302 GCCGCTCAAAATGCCAAATCAGAGATTACCAAGAAATGCGCATGTCTGCTGAAGAACA 361
QY 1892 ttcaaggcagtaagcacctgcccgtcactgtggaagcacacagcagcgttgaggcccc 1951
Db 362 TTCAGGGCATGAAGCAACCTTGGCGGCTCACCTGTGGAAGCACAGCGAGGCTTGGAGGCC 421
QY 1952 tggagaatggaatcaagagctcccggcggtgagaaactctgcagagactttgagctgc 2011
Db 422 TGGAGAATGGAATCAAGAGCTCCCGCGCGCTGGAGAATCTCTGCAGAGACTTTGAGCTGC 481
QY 2012 agaagtggtgttacctaccgctcaacacctctcctcgtcgccactgacccggtcatgc 2071
Db 482 AGAAGGTGTGTTACCTTACCGCTCAACACCTTCTCTCTGGCGGCACTGCACCGGCTCATGC 541
QY 2072 actcaagcagctcctggagcgtgtgcaaacaccaccgccgagccgagccgcagcttca 2131
Db 542 ACTCAAGCAGGTCTCTGGAGCGGCTGTGCAACACCACCGCCGAGCCAGCCGAGCTTCA 601
QY 2132 -gggactcgcgagcgcgtttgacagatcacggagatggtggcacagctccacggtacg 2190
Db 602 GGGGACTGCGAGCGCGCTTTGACAGAGATCACGGAGATGTTGCGCAGCTCCACGCGTACG 661
QY 2191 atgatacagatggagaatttccagaagctgcacgaactcaagaagaatttgatggcatt 2250
Db 662 ATGATCACGATGGAGAATCTCCAGAGCTGCGGAACTCAAGACAAGATTTGATGGCAATT 721
QY 2251 gacaattctgtgtccgggaa- - - -ggaggttcactcctctgggcagcctcagcaaa- - - 2303
Db 722 GACAATCTTGTGGTCCCGGGAAGGGAGTTTCCATCCGGTCTGCGGAGCTTCAAGCAAGC 781
QY 2304 -gctctcgggggaagggtctccagcagcgcatgttcttctctgttcaacagcgtcct-gcta 2361
Db 782 TCCTCGGGGAAAGGACTCCCAAGAGCGGCATGTCTCTCTTGTCAACGAAGTCTCGCTT 841
QY 2362 tacacgagccgggggctgacggcctccaatcagtttaaaagtcacagcgagctcc 2416
|||||
```

EcORI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).\*

Db	842	TACACAGCCGGGCTGACAGGGCTCCATCCACGTTAAAGCCAGGGCAGTCCC	896
RESULT	8		
LOCUS	BF793662		
DEFINITION	602254351f1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4346509 5',	EST	12-JAN-2001
ACCESSION	BF793662		
VERSION	BF793662.1	GI:12098716	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1 (bases 1 to 736)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-re@mail.nih.gov		
	Tissue Procurement: ATCC		
	cDNA Library Preparation: Life Technologies, Inc.		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LLAM9968 row: f column: 14		
	High quality sequence stop: 710.		
FEATURES	Location/Qualifiers		
source	1..736		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4346509"		
	/clone_lib="NIH_MGC_84"		
	/tissue_type="adrenal cortex carcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site:1:		
	NotI; Site:2: SalI; Cloned unidirectionally; oligo-dT		
	primed. Average insert size 1.229 kb. Library enriched for		
	full-length clones and constructed by Life Technologies.		
	Note: this is a NIH_MGC Library."		
BASE COUNT	167 a 228 c 209 g 132 t		
ORIGIN			
	Query Match	22.8%; Score 716; DB 11; Length 736;	
	Best Local Similarity	99.2%; Pred. No. 2.4e-165;	
	Matches 730; Conservative 0; Mismatches 5; Indels 1; Gaps 1;		
QY	2349	cgacgtctctatcacgagccggggctgacggcctccaatcagtttaaaagtcacgg	2408
Db	1	CGACGTCCTGCTATACACGAGCGGGGCTGACGGCTCCATCATCTTAAGTCCACGG	60
QY	2409	gcagctcccgctctatgcatgacgattgagagagcgaagcagtgagggggtgccca	2468
Db	61	GCAGTCCCGCTCTATGGCATGACGATTGAGGAGAGCGAGTGGGGGTGCCCA	120
QY	2469	ctgcctgacctccgggagcagtcgagtcgagtcgagtcgagtcgagtcgagtcgag	2528
Db	121	CTGCGCTGACCCCTCCGGGGGACGCGGAGTCCATCATCTGTTGGCGCGGCTCGG	180
QY	2529	gatgagaagtgggtgagagatccagatgcccattgacctggcgagagagcagcag	2588
Db	181	GATGGAGAAGTGGTGTGAGACATCCAGATGGCCATTGACCTGGCGGAGAGAGCAG	240
QY	2589	ccccgccctgagttctctgcccagcagccccctcgacaaagtcacctgatgaagccac	2648
Db	241	CCCGGCCCTGAGTCTCTGCGCCAGCAGCCCCCTGTGACACAAAGTCCCTGATGA	300
QY	2649	cgcggtgacccaggagtcagagatgacctgagcgcctcgccacatcgtggagcgcca	2708
Db	301	CGGGGTGACAGGAGTACAGGATGACCTGAGCGCTCGCGGCATCGCTGAGCGCCA	360
QY	2709	ggcccgccgacccgagcgaacacaaatggtgacagtggtgacagtcgagcagcagtcctc	2768
Db	361	GGCCCGCACCGCGGCAACACAATGTTGACAGTGTGTGGCACCAGCAACACCGCTCTC	420
QY	2769	catgtggacttcagcatcgagtcagtcagtcagtcagtcagtcagtcagtcagtcagtc	2828
Db	421	CATGTTGGGCTTTCAGCATGCTGAGGAAATCAGTTGCTTGGAAACCTGCTGAGGAAAT	480
QY	2829	caaaaacagacgagtggtgagcagagctggtggtggtggtggtggtggtggtggtggt	2888
Db	481	CAAAAACAGACAGCGGTGGCAGAGCTGTGGGTGGTGTTCACAAAATTCCTGCTTCTT	540
QY	2889	ctacaaatcacaccagagacaatcatcccttggcagcgcctcctcgtcgtcgtcgtcgtc	2948
Db	541	CTACAAATCACACAGGACAATCATCCCTTGGCAGCTGCTGCTGCTGCTGCTGCTGCT	600
QY	2949	caccatccctctgagtcgagaaacatccagaaagactcagtcagtcagtcagtcagtc	3008
Db	601	CACCATCCCTCTGAGTCCGAGAACATCCAGAAAGACTAGGTGTTCAAGCTGCACTTCA	660
QY	3009	gtccacgtctactacttcagggcgagagcagtcacagtcacagtcacagtcacagtc	3067
Db	661	GTCCACGCTCTACTACTTTCAGGGCGGAGAGAGTTCAGCTTCGAAAGGTGGTGGCA	720
QY	3068	tgatccgagtcgac	3067
Db	721	TGAATCCGAGTGCAC 736	
RESULT	9		
LOCUS	BE745887		
DEFINITION	601573513f1 NIH_MGC_9 Homo sapiens cDNA clone IMAGE:3834699 5',	EST	15-SEP-2000
ACCESSION	BE745887		
VERSION	BE745887.1	GI:10159879	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	1 (bases 1 to 898)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs-re@mail.nih.gov		
	Tissue Procurement: DCTD/DRP		
	cDNA Library Preparation: Ling Hong/Rubin Laboratory		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov		
	Plate: LLCM515 row: i column: 04		
	High quality sequence stop: 723.		
FEATURES	Location/Qualifiers		
source	1..898		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:3834699"		
	/clone_lib="NIH_MGC_9"		
	/tissue_type="adenocarcinoma cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: ovary; Vector: pOTB7; Site:1: XhoI; Site:2:		
	EcoRI; cDNA made by oligo-dT priming. Directionally		
	cloned into EcoRI/XhoI sites using the following 5'		
	adaptor: GGCACGAG(G). Size-selected >500bp for average		
	insert size 1.8kb. Library constructed by Ling Hong in		
	the laboratory of Gerald M. Rubin (University of		
	California, Berkeley) using ZAP-cDNA synthesis kit		
	(Stratagene) and Superscript II RT (Life Technologies)."		

```

BASE COUNT      236 a      229 c      269 g      164 t
ORIGIN

Query Match      22.5%; Score 705.4; DB 10; Length 898;
Best Local Similarity 96.8%; Pred. No. 1e-162;
Matches 762; Conservative 0; Mismatches 21; Indels 4; Gaps 4;

QY 613 ggcaaacaccagcaagaatcagatttccagctcctagagattccgctgctgagagatg 672
DB 1 GGACAAACACACAGCAGATCAGATTTCCAGCTCCTAGAGATTCCGCTCGCTAGAGATG 60
QY 673 tatggaatccggttgaccgcggcgaagagcaggaagggcaggaagatcaatctggccgtt 732
DB 61 TATGGAATCCGTTGACCCCGCCCAAGGACAGAGGACAGAGATCAATCTGCGCGTT 120
QY 733 gcaaacacgggaattcttagtttcaggtttcactaagaatcaatgccttcaactgggcc 792
DB 121 GCCAACACGGGAATCTAGTGTTCAGGGTTTCACTAAGATCAATGCTTCAACTGGGCC 180
QY 793 aaggtcggaagctgagcttcaagagaaagcgccttctcatcaagctccgcccagatgcc 852
DB 181 AAGTGGGAAGCTGAGCTTCAAGAGAAAGCGCTTCTCATCAAGCTCCGGCCAGATGCC 240
QY 853 aatagtcgtaccagataaccttggaaattcctgatggcagtcgggatttcttcaagtcc 912
DB 241 AATAGTGCGTACCAAGATACCTTTGGAATTCCTGATGCCAGTCGGGATTTCTGCAAGTCC 300
QY 913 ttctggaatactgtgttgaacatcatgccttcttttagactttttgaagagcccaacca 972
DB 301 TTTGTGAAAATCTGTG-TGAACATCATGCTCTTCTTAGACTTTTGAAGAGGCCCAACCA 359
QY 973 aagcccaagccctcctttagccgggggtatcatttgcgttcagtggtcgactcag 1032
DB 360 AAGCCCAAGCCCTCTCTTTAGCCGGGGTCAATATTCGGTTCAAGTGTGCGACTCAG 419
QY 1033 aagcaggttctcactatgttaa-agaagagagacataagaaggtgcagtttgaaggaa 1091
DB 420 AAGCAGGTTCTCGACTATGTGTAACAGAAGAGGAGACATAAGAAGGTGCAGTTTGAAGGAA 479
QY 1092 gcacagaagattcattctatccgagccttgcttccacagcctacagaactgaattcgga 1151
DB 480 GCACAGAAGATTCATTCTATCCGAGCCCTTGCTTCCACAGCCTACAGAACTGAATTCGGA 539
QY 1152 agtgcgtgagcagctcagcagagcaccagccttaccatttggagaaggtgcgaactccc 1211
DB 540 AGTGCTGGACCACTCTCAGCAGAGACCAGCCTTACATTTGGAGAGAGGTGCCAATCTCC 599
QY 1212 agggggccagagctgcgaggaaggaagcaggaaggtttccgcccgggagccgggtc 1271
DB 600 A-GGGGCCAGAGCTGCGCGGAGGAAAGGAAACGAAAGGTTTCCGCGGGGAGCCGGGTC 658
QY 1272 gcaccagccttcgcccagagagagcccggtgtaacaagcagcgagcagccgc 1331
DB 659 GCACCCAGAGCCCTGCGCGGAGGAGAAAGCCCGGGGTAAACAGCAGCGGAGCGCCG - 717
QY 1332 ctgagcgcacagaggaagagagaggtcggttaaggaTaggaccacagcagagtaaac 1391
DB 718 CTCGGCGCCACCGAGGAGAGAGAGGAGGCGCTTAAGATAGGACCCACAGAGAGTAAAC 777
QY 1392 tcagccc 1398
DB 778 TTAGCCC 784

RESULT 10
BG829192
LOCUS
DEFINITION 602753246F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4906114 5',
mRNA sequence.
ACCESSION BG829192
VERSION BG829192.1 GI:14176766
KEYWORDS EST.
```

```

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 835)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llni.gov
Plate: LLCM1808 row: k column: 11
High quality sequence stop: 835.
FEATURES
Location/Qualifiers
1..835
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NIH_MGC_17"
/tissue_type="rhabdomyosarcoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: muscle; Vector: pORF7; Site_1: EcoRI;
Site_2: XhoI; cDNA made by oligo-dr priming.
Directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life technologies)."
```

```

BASE COUNT      213 a      231 c      243 g      148 t
ORIGIN

Query Match      22.5%; Score 704.6; DB 11; Length 835;
Best Local Similarity 95.8%; Pred. No. 1.6e-162;
Matches 800; Conservative 0; Mismatches 24; Indels 11; Gaps 7;

QY 928 gttgaacatcatgcttctttagacttttgaagagcccaaacagccagccgcgc 987
DB 2 GTTGAACATCATGCTTCTTTA-ACITTTTGAAGAGCCCAACCAAGCCCAAGCCGTC 60
QY 988 ctctttagccgggggtcatttcggttcagtcggttcgactcagaagaggtcttcgac 1047
DB 61 CTCTTTAGCCGGGGTCAATCATTTTCGTTTCAGTGTGCGACTCAGAAGCAGGTCTCGAC 120
QY 1048 tatgttaagaagagagacataagaaggtgcagtttgaaggaagcacagcaagattcat 1107
DB 121 TATGTTAAAGAGAGGAGACATAAGAAGGTGCAGTTTGAAGAGGAGCAGCAAGATTCAT 180
QY 1108 tctatccgagccttcttcacagcctacagaactgaattcggaagtgcgtggagcagctc 1167
DB 181 TCTATCCGGAGCCTTCTTCACAGCCTACAGAACTGAATTCGGAAGTGTCTGGAGCAGTCT 240
QY 1168 cagcagagcaccagccttaccatttgagaaggtgcgcgaatctccagggggccagagctgc 1227
DB 241 CAGCAGAGCACCAGCCTTACATTTGGAGAAGGTGCCGAATCTCCAGGGGGCCAGAGCTGC 300
QY 1228 cggcaggaagaacccgaaggtttccgcggggagccggggtcgaccaccgagccctgcg 1287
DB 301 CGCGAGGAAGAAACCGAAGGTTTCCGCGGGAGCCCGGGGTCCGACCCAGCCCTCGC 360
QY 1288 ccgagagagcccccggggtaacaagcagggcgagcagccctctcgccgccacaggag 1347
DB 361 CCGAGGAGAAGCCCCCGGGTAAACAAGCAGCGGACGAGCGCCCTCGGCGCCACCGAG 420
QY 1348 gaagagagagaggtcgtt-aaggataggagccagagagtaaaccttcagcccccgagcc 1406
|||||
```









```
QY 1217 gccagagctgcgcgaggaagaaacgaaggtttccgcgggagcggggtgcacc 1276
Db 120 GCCAGAGTGCcgcgcgaggaagaaacgaaggtttccgcgggagcggggtgcacc 179
QY 1277 cgagccctgcgcgcgaggaagccccgcgggtaacaagcagcggagcgcgcctcgg 1336
Db 180 CGAGCCCTGCcgcgcgaggaagccccgcgggtaacaagcagcggagcgcgcctcgg 239
QY 1337 cgcacacgaggaagagagaggtcgtaagga taggacccagcagagtaaacctcagc 1396
Db 240 CGCCACGAGGAGAGAGAGAGGTCGTAAAGTAGAGACCCAGCAGAGTAACCTCAGC 299
QY 1397 cccgcagcacaagcagcctcctgactgcagtcgactccactcccttcgcagctgtgga 1456
Db 300 CCCCAGCAGCAGCAGCAGGTCCTGACTGGCAGTCCTACCTTTCCGAGCTGTCTGTGA 359
QY 1457 actcaggggggagtgccctgcacacgtgaccttgtctcccaacctgagccccgcaca 1516
Db 360 ACTCGAGGGGGAGTGGCCCTGCGCAACGTGACCTTGTCTCCCAACCTGAGCCCGACA 419
QY 1517 coaagcagcctctccttgatcagccgcgtgctgaatgacacgacctgccccgcaggg 1576
Db 420 CCAAGCAGCGCTCTCCCTTGATCATCCCGCTGCTGAATGACCAAGG-CTGCCCCCGAGCG 478
QY 1577 acgatgagatgagggcgggaggaagagattcccaactgataaagcgtacttcatagcta 1636
Db 479 AGGATGAGGATGAGGGCGGAGGAGAGATTCCCAACTGATAAAGCGTACTTCATAGCTA 538
QY 1637 aggaagtgtctaccacccagcgaacatatctgaaggtctcgaagttatcacttcgtggt 1696
Db 539 AGGAAGTGTCTACCAACCGAGCAATATCTGGAAGATCTCGAAGTTATCATCTTCTGGT 598
QY 1697 ttcagagcacagtgcgaagagagcgcacatgccggaagcactgaaagtctcatattcc 1756
Db 599 TTCAGAGCACAGTAGCAAAAGAGAGGCGCATGCCGGAAGCACTGAAAGTCTCATAGTCC 658
QY 1757 cgaatttgaaaccttgacaaattctactaataattctcgaagaaattgcagaaacagac 1816
Db 659 CGAATTTGGAACCTTGGACAAATTTTCATATAATTTTCTCCA--GGAAATTGAGCAACGAC 716
QY 1817 ttgcctgtgggaagcgcgtcaaatgcccaatcacagaga 1856
Db 717 TTG-CCTGTGGGAAGCGCTCAAAATGCCAAATCAGAGATA 755

RESULT 14
LOCUS BG767698 891 bp mRNA EST 15-MAY-2001
DEFINITION 602741690F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4871374 5',
mRNA sequence.
ACCESSION BG767698
VERSION BG767698.1 GI:14078351
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 891)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC/DCTD/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI746 row: c column: 23
High quality sequence stop: 811.
Location/Qualifiers
```

```
source 1. 891
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4871374"
/tissue_type="melanotic melanoma, high MDR (cell line)"
/lab_host="DH10B (phage-resistant)"
/organism="Homo sapiens"
/vector="pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCACGAG(G). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library. !"
BASE COUNT 248 a 216 c 249 g 177 t 1 others
ORIGIN
Query Match 21.6%; Score 678.2; DB 11; Length 891;
Best Local Similarity 94.2%; Pred. No. 5e-156;
Matches 801; Conservative 0; Mismatches 39; Indels 10; Gaps 9;
QY 436 caaggcaggttgacgtgtaatagacaccag-cgcagctctcttgatttcacacattgtgca 494
Db 2 CAAGCAGGTTGACGTGTAATGACACACGACGACGCTCTCTTTGATTTCACACA-TGTGCA 60
QY 495 atctgagatggggtatttgatgaagccttgacagagacgttagcaaaaaataaata 554
Db 61 ATCTGAGAT--GGGATTTGATGAAGCCCTTGGCAGAGAGACACTTAGCAAAAAATAAATA 118
QY 555 catacctcagcaagcgcactagagacaaaatctggaatttcacacataaacacattgg 614
Db 119 CATACCTCAGCAAGACGCACTAGAGACAAAAATCGTGGAAATTTCAACATAACACA-TGG 177
QY 615 acaaacaccagcagacatcagatttcagctcctagatgctgcgcgtcggtcagagatgta 674
Db 178 ACAAAACACAGCAGAAATCAGATTTCAGCTCCTAGAGATTGCCCGTCGGCTAGAGATGTA 237
QY 675 tggaaatccggttgcccccgcgaagacaggggaagcagcagagatcaatctgcccgttgc 734
Db 238 TGGAAATCCGGTTGCACCCGCCCAAGCAGGAGGACAGAGATCAATCTGCGCGTTGC 297
QY 735 caacacgggaattctagtgtttcagggtttcactaagatcaatgccttcaacttggcccaa 794
Db 298 CAACACGGGAATCTAGTGTTCAGGGTTTCTACTAAGATCAATGCTTCAACTGGGCCAA 357
QY 795 ggtcgggaagctgagcttcaaggaagcgtttctcatcaagctccgcccagatgcaca 854
Db 358 GGTGCGGAAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGCTCCGCCAGATGCCAA 417
QY 855 tagtcgtaccaggatcaccttggaattcctcgtatgcccagtcgggaattctgcaagtcctt 914
Db 418 TAGTCCGTACCAAGGATACCTTGGAAATCTCTGATGCCAGTCGGGATTTCTCAAGTCCCT 477
QY 915 ctggaataatctgtgtgaacatcatgcctctttagacttttgaagagcccaaaccaaa 974
Db 478 CTGGAATAATCTGTGTGAACATCATGCCCTTCTTAGACTTTTGAAGAGCCCAACCAAA 537
QY 975 gcccaagccgctcctcttagcgggggtcatcatttcagttcagtggtcggactcagaa 1034
Db 538 GCCCAAGCCCGCTCTTTAGCCGGGGTTCATCATTTTCGGTTTCAGTGGTCGAGACTCAGAA 597
QY 1035 gcagttctcagactatgtt-aaagaagaggagacataagaaggtgcagtttgaaggaagc 1093
Db 598 GCAGGTTCTCGACTATGTTAAAGAAGAGGAGGACATAAGAAGGTGCAG-TTGAAGGAAGC 656
QY 1094 acagcaagattcattctatcccgagccttgcgttcacagcgtcagaaactgaattcgg-aa 1152
Db 657 ACAGCAAGATTTCATTCATCCGAGCGCTTGTGCGACAGCTACAGAACTGAATTCGGAAA 716
QY 1153 gtgtgagagcagtcctcagcagagcaccaccttacctatttgagagaaggttcggaattccaa 1212
```



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 6, 2001, 08:51:28 ; Search time 283.82 Seconds  
(without alignments)  
10397.130 Million cell updates/sec

Title: US-09-555-342A-1

Perfect score: 3442

Sequence: 1 cgccgcagccgcgcgcgtg.....agtattaaacattgtcatt 3442

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3442	100.0	3442	20 AAX79183	Human chondrocyte-
2	3437.2	99.9	4687	21 AAA08582	Human cytoskeleton
3	1688.6	49.1	3094	21 AAC98992	Human pancreatic c
4	223.6	6.5	241	21 AAA42150	Human secreted exp
5	169.6	4.9	3620	21 AAA08581	Human cytoskeleton
6	151.6	4.4	544	22 AAH98023	Murine 7-transmemb
7	144	4.2	3984	18 AAT58627	Protein tyrosine p
8	144	4.2	3984	20 AAX78463	Human PTPH1 CDNA
9	143.2	4.2	2872	22 AAS08887	Human protein tyro
10	133.6	3.9	3166	22 AAH17787	Human cDNA sequenc
11	95.8	2.8	659	22 AAH33845	Human colon cancer

12	85	2.5	665	22 AAH07867	Human cDNA clone (
13	85	2.5	1821	22 AAH15014	Human cDNA sequenc
14	76	2.2	410	21 AAC69420	Human secreted pro
15	69.4	2.0	1156	22 AAH99833	Human protein enco
16	63.8	1.9	1626	22 AAI59140	Human polynucleoti
17	63.8	1.9	2556	22 AAI60926	Human polynucleoti
18	62	1.8	1860	22 AAF99965	Human colon carcin
19	62	1.8	5923	22 AAS02052	DNA encoding molec
20	62	1.8	5923	22 AAS02052	DNA encoding molec
21	59.2	1.7	398	21 AAH30288	Human colon cancer
22	58.4	1.7	2355	22 AAH99674	Human protein enco
23	56.8	1.7	4080	19 AAV34368	Human protein tyro
24	56.8	1.7	4080	20 AAZ11221	Human protein tyro
25	56.2	1.6	365	14 AAQ39923	Expressed Sequence
26	56.2	1.6	365	14 AAQ39935	Human brain expres
27	54.4	1.6	1778	22 AAI60338	Human polynucleoti
28	54	1.6	550	22 AAH06244	Human cDNA clone (
29	52.8	1.5	556	22 AAH99189	Human protein enco
30	50.2	1.5	3115	21 AAC98113	Human colon cancer
31	48.2	1.4	454	22 AAI15794	Probe #5727 for ge
32	48.2	1.4	454	22 AAI37576	Probe #6262 used t
33	48.2	1.4	1730	22 AAI62819	Human cDNA SEQ ID
34	48.2	1.4	1816	22 AAI58552	Human polynucleoti
35	47.4	1.4	2400	17 AAT42859	Choline oxidase ge
36	47.2	1.4	160	22 AAI25012	Probe #14945 for g
37	47.2	1.4	160	22 AAI50750	Probe #19436 used
38	46.6	1.4	1696	22 AAH16459	Human cDNA sequenc
39	45.8	1.3	2400	18 AAT75000	Choline oxidase ge
40	44.6	1.3	347	21 AAC02360	Human secreted pro
41	44.6	1.3	625	22 AAH07308	Human cDNA clone (
42	44.4	1.3	342	21 AAF08796	Fusarium venenatum
43	44.2	1.3	10732	21 AAA10594	Gene encoding a su
44	44	1.3	463	15 AAQ68827	hgt4 clone showing
45	42	1.2	4610	21 AAA27896	Human GEF containi

#### ALIGNMENTS

#### RESULT 1

AAX79183  
ID AAX79183 standard; DNA; 3442 BP.  
AC AAX79183;  
XX  
XX  
DT 17-AUG-1999 (first entry)  
XX  
DE Human chondrocyte-derived gene CDEP.  
XX  
KW Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;  
KW Db1 homology domain; pleckstrin homology domain; rheumatoid 'arthritis;  
KW drug; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO928458-A1.  
XX  
XX  
PD 10-JUN-1999.  
XX  
PF 27-NOV-1998; 98WO-JP05348.  
XX  
XX 27-NOV-1997; 97JP-0342060.  
PR (CHUS ) CHUGAI SEIYAKU KK.  
XX  
PI Kato Y, Kawamoto T, Koyano Y;  
XX  
XX WPI; 1999-371117/31.  
DR P-PSDB; AAY07482.  
XX  
XX Protein CDEP expressed in differentiated chondrocytes, and gene  
PT encoding it  
XX

PS Claim 5; Fig 1; 59pp; Japanese.

CC This sequence represents the coding region for a protein (CDEP) expressed in differentiated human foetal chondrocytes, which contains an exrin-like domain, a Dbl homology (DH) domain and a pleckstrin homology (PH) domain. The nucleic acid or protein can be used in the investigation and treatment of cancers and arthritic diseases (including chronic rheumatoid arthritis), or for screening of candidate anticancer drugs.

XX Sequence 3442 BP; 864 A; 952 C; 927 G; 699 T; 0 other;

Query Match 100.0%; Score 3442; DB 20; Length 3442;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 3442; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	cgccgagccgcccgcgctgtggagatattcttaacgcttctcatcattgaggagaata	60
DB	1	cgccgagccgcccgcgctgtggagatattcttaacgcttctcatcattgaggagaata	60
QY	61	gagcagagccgacccagagatcacgactggggcccggaattcgggatcagtacc	120
DB	61	gagcagagccgacccagagatcacgactggggcccggaattcgggatcagtacc	120
QY	121	ttggaacgtggacagaagccgcccacacacttcaggaatactcgtgtccatcaaatc	180
DB	121	ttggaacgtggacagaagccgcccacacacttcaggaatactcgtgtccatcaaatc	180
QY	181	cagatcgtggatcacacccagaggagcatttgaagtcccaaaagatcctctgggaaggtg	240
DB	181	cagatcgtggatcacacccagaggagcatttgaagtcccaaaagatcctctgggaaggtg	240
QY	241	ctgctgatgcagtttgaacacactcaactcgtggaagtgactattttggcctcgag	300
DB	241	ctgctgatgcagtttgaacacactcaactcgtggaagtgactattttggcctcgag	300
QY	301	tttctgatcacaaaagatacacggtgtggtggtatctctaaacccattgtgaaacag	360
DB	301	tttctgatcacaaaagatacacggtgtggtggtatctctaaacccattgtgaaacag	360
QY	361	attagaagccaaagacggtgtgttgaattgtgtggtgaaattcttcgcctgaccac	420
DB	361	attagaagccaaagacggtgtgttgaattgtgtggtgaaattcttcgcctgaccac	420
QY	421	acaaactccaaagaagactcacaggtacctgttcgcgtgcaggtgaagcagacttg	480
DB	421	acaaactccaaagaagactcacaggtacctgttcgcgtgcaggtgaagcagacttg	480
QY	481	gctcaagggcaggttgacgtgtaatacacacagcgcagctctctgtatttcacacattgtg	540
DB	481	gctcaagggcaggttgacgtgtaatacacacagcgcagctctctgtatttcacacattgtg	540
QY	541	caactcagcagcagcactagagacaaactcgtggaatttcaccataaacacatt	600
DB	541	caactcagcagcagcactagagacaaactcgtggaatttcaccataaacacatt	600
QY	601	tacatacctcagcagcagcactagagacaaactcgtggaatttcaccataaacacatt	660
DB	601	tacatacctcagcagcagcactagagacaaactcgtggaatttcaccataaacacatt	660
QY	661	ggacaaacacacagcagaatacagatttcacgctcctagagattgcccctcggttagagatg	720
DB	661	ggacaaacacacagcagaatacagatttcacgctcctagagattgcccctcggttagagatg	720
QY	721	tatggaatccggttgcacccgccaagacagggaggaagcaggaatcaatctggcgtt	780
DB	721	tatggaatccggttgcacccgccaagacagggaggaagcaggaatcaatctggcgtt	780
QY	781	gccaacacgggaaattctagttttcaggttttcaactaagatcattcctcaactgggcc	840
DB	781	gccaacacgggaaattctagttttcaggttttcaactaagatcattcctcaactgggcc	840
QY	841	aaggtgcggaagctgagcttcaagagagaagcgttttctcatcaagctccggccagatgcc	900

DB	841	aaggtgcggaagctgagcttcaagagagaagcgtttctcatcaagctccggccagatgcc	900
QY	901	aatagtcgtaccaggaatacttggaattctctgatgcccagtcgggatttctgcaagtcc	960
DB	901	aatagtcgtaccaggaatacttggaattctctgatgcccagtcgggatttctgcaagtcc	960
QY	961	ttctggaaaatctgtgtgaaacatcatgcctctttagactttttgaagagcccaaacca	1020
DB	961	ttctggaaaatctgtgtgaaacatcatgcctctttagactttttgaagagcccaaacca	1020
QY	1021	aagcccaagccgctctctttagccgggggtcatcattcgggttcagtggtcggaactcag	1080
DB	1021	aagcccaagccgctctctttagccgggggtcatcattcgggttcagtggtcggaactcag	1080
QY	1081	aagcaggtttctgaactatgttaaagaagaggagacataagaaggtgcagtttgaagaag	1140
DB	1081	aagcaggtttctgaactatgttaaagaagaggagacataagaaggtgcagtttgaagaag	1140
QY	1141	cacagaagattcattctatccgagccttgccttcacagctcacagaactgaattcgaa	1200
DB	1141	cacagaagattcattctatccgagccttgccttcacagctcacagaactgaattcgaa	1200
QY	1201	gtcgtggagcagttctcagcagagcaccagccttacatttggaaggtgcggaattctcca	1260
DB	1201	gtcgtggagcagttctcagcagagcaccagccttacatttggaaggtgcggaattctcca	1260
QY	1261	gggggccaagctgcgcggcaggaagaaacccgaaggtttccgcggggagccgggctg	1320
DB	1261	gggggccaagctgcgcggcaggaagaaacccgaaggtttccgcggggagccgggctg	1320
QY	1321	caaccagcctgcgcggcaggaagcccgccggttaacaagcagcgagcgagcgccg	1380
DB	1321	caaccagcctgcgcggcaggaagcccgccggttaacaagcagcgagcgagcgccg	1380
QY	1381	tcgggcccacgaggaagaggaggttcgttaaggataggagcccgagagtaaacct	1440
DB	1381	tcgggcccacgaggaagaggaggttcgttaaggataggagcccgagagtaaacct	1440
QY	1441	cagcccccagcagcaaacacagctcctcactgagtcactcacttccagactctct	1500
DB	1441	cagcccccagcagcaaacacagctcctcactgagtcactcacttccagactctct	1500
QY	1501	gtgaactcgcaggggggagtgcccttcgcaacgtgacctgtctcccaacctgagcccc	1560
DB	1501	gtgaactcgcaggggggagtgcccttcgcaacgtgacctgtctcccaacctgagcccc	1560
QY	1561	gacacaaagcagggcctctccttgatcagcccctgtgtaatacagaagcctgcccccg	1620
DB	1561	gacacaaagcagggcctctccttgatcagcccctgtgtaatacagaagcctgcccccg	1620
QY	1621	acggacgatgagatgagggccgaggaagagattcccaactgataaagcgtacttcata	1680
DB	1621	acggacgatgagatgagggccgaggaagagattcccaactgataaagcgtacttcata	1680
QY	1681	gctaagaaagtgtctaccacgcagcagacatactgaaggtctcgaagtctcacttcg	1740
DB	1681	gctaagaaagtgtctaccacgcagcagacatactgaaggtctcgaagtctcacttcg	1740
QY	1741	tggtttcagagcagctgagcaaaagagacgccaatgcgggaagcactgaaaaagtctcata	1800
DB	1741	tggtttcagagcagctgagcaaaagagacgccaatgcgggaagcactgaaaaagtctcata	1800
QY	1801	ttcccgaatttgaaccttgcacaaatttcatactactaatttttccaaggaaattgagcaa	1860
DB	1801	ttcccgaatttgaaccttgcacaaatttcatactactaatttttccaaggaaattgagcaa	1860
QY	1861	cgactgcctgtgggaagccgctcaaatgcacaaatcagagattaccacaaagtggcc	1920
DB	1861	cgactgcctgtgggaagccgctcaaatgcacaaatcagagattaccacaaagtggcc	1920
QY	1921	gagtcatgctgaagaacacattcaggggcatgaagcactgcgcgtcactctgtggaagcac	1980

Db 1921 gatgtcatgctgaagaacattcagggcacatgaagcaccctggcggctcacctgtggaagcac 1980  
QY 1981 agcagagccttgaggccctggagaatgaatacaagatcccgcggtggagaacttc 2040  
Db 1981 agcagagccttgaggccctggagaatgaatacaagatcccgcggtggagaacttc 2040  
QY 2041 tgcagagacttgagctgcagaaggtgtgttacctaccgctcaaaccttctccctgcgg 2100  
Db 2041 tgcagagacttgagctgcagaaggtgtgttacctaccgctcaaaccttctccctgcgg 2100  
QY 2101 ccaactgcacggctcatgcactacaagcaggttctctgagcggctgtgcaaaacacaccccg 2160  
Db 2101 ccaactgcacggctcatgcactacaagcaggttctctgagcggctgtgcaaaacacaccccg 2160  
QY 2161 ccagccacgcgcgaacttcaggactcccgagccgcttggcagagatcacggagatgggtg 2220  
Db 2161 ccagccacgcgcgaacttcaggactcccgagccgcttggcagagatcacggagatgggtg 2220  
QY 2221 gcacagctccacgggtacgatgatcaagatggagaatttcagaagctgcacgaactcaag 2280  
Db 2221 gcacagctccacgggtacgatgatcaagatggagaatttcagaagctgcacgaactcaag 2280  
QY 2281 aaagatttgattggcattgacaatctgtgttccgggaaggagttcatccgtctgggnc 2340  
Db 2281 aaagatttgattggcattgacaatctgtgttccgggaaggagttcatccgtctgggnc 2340  
QY 2341 agcctcagcaagctctcggggaagggtccacgacgcatgttcttctctgttcaacgac 2400  
Db 2341 agcctcagcaagctctcggggaagggtccacgacgcatgttcttctctgttcaacgac 2400  
QY 2401 gtctgtctatacagcagcgggggttgacggcctcaaatcagtttaaaagtccaacggcag 2460  
Db 2401 gtctgtctatacagcagcgggggttgacggcctcaaatcagtttaaaagtccaacggcag 2460  
QY 2461 ctcccgctctatgcatgacgtatgagagcagcagagcagagtggtgggtgcccactgc 2520  
Db 2461 ctcccgctctatgcatgacgtatgagagcagcagagcagagtggtgggtgcccactgc 2520  
QY 2521 ctgacctccggggccagcggcagtcacatcatctgtgcccgcagttctcgttcgagatg 2580  
Db 2521 ctgacctccggggccagcggcagtcacatcatctgtgcccgcagttctcgttcgagatg 2580  
QY 2581 gagaagtgggttgagagatccaagatggccattgacctggcgagagaagcagcagcccc 2640  
Db 2581 gagaagtgggttgagagatccaagatggccattgacctggcgagagaagcagcagcccc 2640  
QY 2641 gcccttgattctctgccagcagccccctgacaacaagtcacctgatgaagccaccgag 2700  
Db 2641 gcccttgattctctgccagcagccccctgacaacaagtcacctgatgaagccaccgag 2700  
QY 2701 gctgaccagagtcagagatgacctgagcgcctcgcacatcgtctggagcgcagggcc 2760  
Db 2701 gctgaccagagtcagagatgacctgagcgcctcgcacatcgtctggagcgcagggcc 2760  
QY 2761 ccgcacccggcgaacaacatgtgtcacgtgtgtgtgcaccgcgaacaccagcgtctccatg 2820  
Db 2761 ccgcacccggcgaacaacatgtgtcacgtgtgtgtgcaccgcgaacaccagcgtctccatg 2820  
QY 2821 gtggaacttcagatcgcagtggaagaatcagttgtctggaaacctgctgaggaattcaaa 2880  
Db 2821 gtggaacttcagatcgcagtggaagaatcagttgtctggaaacctgctgaggaattcaaa 2880  
QY 2881 aacagcaacgggtggcagaagctgtgggtgtgttcacaaacttgcctgttcttctac 2940  
Db 2881 aacagcaacgggtggcagaagctgtgggtgtgtgttcacaaacttgcctgttcttctac 2940  
QY 2941 aaatcacaccaggacaatacatcccttgcagcgtcctctgctcgtcgtactgcctcaac 3000  
Db 2941 aaatcacaccaggacaatacatcccttgcagcgtcctctgctcgtcgtactgcctcaac 3000  
QY 3001 atccccctctgagtcgcagaaatccagaaagactacgtgttcaagctgcacttcaagtcc 3060  
Db 3001 atccccctctgagtcgcagaaatccagaaagactacgtgttcaagctgcacttcaagtcc 3060

QY 3061 caccgtactacttcaggcgggaaagcggagtacacgttcgaaaaggtggatggaagtgtac 3120  
Db 3061 caccgtactacttcaggcgggaaagcggagtacacgttcgaaaaggtggatggaagtgtac 3120  
QY 3121 ccagtgccacacagctctcctcgcgcacccacgctgttgagccacaaagagctctctgtg 3180  
Db 3121 ccagtgccacacagctctcctcgcgcacccacgctgttgagccacaaagagctctctgtg 3180  
QY 3181 tattgatggccggacacactcgtttccgcagtggtgcttccctggaagacgtttccctt 3240  
Db 3181 tattgatggccggacacactcgtttccgcagtggtgcttccctggaagacgtttccctt 3240  
QY 3241 cttctgtattaatgaagcctggttaaaattaaacacctgtctgaaaaatacaaaacatggctt 3300  
Db 3241 cttctgtattaatgaagcctggttaaaattaaacacctgtctgaaaaatacaaaacatggctt 3300  
QY 3301 ccacgacgctctcctgtctccacagcgcggtttttaaccccgcacctctcagcgtttgaa 3360  
Db 3301 ccacgacgctctcctgtctccacagcgcggtttttaaccccgcacctctcagcgtttgaa 3360  
QY 3361 tgaacagcgtcccccacctccagtcctcctggcatccgctggggcgctgttctttagctagt 3420  
Db 3361 tgaacagcgtcccccacctccagtcctcctggcatccgctggggcgctgttctttagctagt 3420  
QY 3421 ccagttataaaacattgtcatt 3442  
Db 3421 ccagttataaaacattgtcatt 3442

## RESULT 2

AAA08582

ID AAA08582 standard; DNA: 4687 BP.

XX AC AAA08582;

XX DT 19-JUL-2000 (first entry)

XX DE Human cytoskeleton associated protein 2 (CYSKP-2) coding sequence.

XX KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;

XX KW autoimmunity; inflammatory; vesicle trafficking; neurological;

XX KW cardiovascular; cell motility; reproductive; muscle disorder; ss.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT CDS 90..3227

XX FT /\*tag= a

XX FT /product= CYSKP-2

XX PN WO200017355-A2.

XX PD 30-MAR-2000.

XX PF 17-SEP-1999; 99WO-US21565.

XX PR 18-SEP-1998; 98US-0172226.

XX PR 27-APR-1999; 99US-0131321.

XX PA (INCY-) INCYTE PHARM INC.

XX PI Lal P, Tang YT, Yue H, Hillman JL, Bandman O, Corley NC;

XX PI Guegler KJ, Patterson C, Azimzai Y, Baughn MR;

XX DR WPI: 2000-283582/24.

XX DR P-PSDB; AAY91947.

XX XX Human cytoskeleton associated proteins, used to treat cell

XX PT proliferative, autoimmune/inflammatory, vesicle trafficking,

XX PT neurological, cell motility, reproductive and muscle disorders

XX PS Claim 9; Page 101-102; 113pp; English.

```

xx AAA0581-96 encode human cytoskeleton associated proteins 1 to 16
CC (CYSP-1 to CYSP-16) respectively. The sequences can be used to treat
CC and diagnose cancer and cell proliferative, autoimmune/inflammatory,
CC vesicle trafficking, neurological, cardiovascular, cell motility,
CC reproductive and muscle disorders. Pharmaceutical compositions
CC containing CYSP-2 to CYSP-16 can be used to treat or prevent disorders
CC associated with decreased expression or activity of CYSP (claimed), for
CC example, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,
CC cancers, autoimmune/anti-inflammatory disorders such as allergies, anemia,
CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,
CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma
CC and trauma. CYSP antagonists can be used to treat or prevent a
CC disorder associated with increased expression or activity of CYSP
CC (claimed).
xx
SQ Sequence 4687 BP; 1179 A; 1261 C; 1196 G; 1051 T; 0 other;

Query Match          99.9%; Score 3437.2; DB 21; Length 4687;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 3439; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1  cgcgcgagccgcgcgtgagatattcttaagccgtttcatcattggaagaata 60
Db 42  cgcgcgagccgcgcgtgagatattcttaagccgtttcatcattggaagaata 101

Qy 61  gacagagccgacccaggaacacgactggggcccgaaaaattcgggagcagtaac 120
Db 102 gacagagccgacccaggaacacgactggggcccgaaaaattcgggagcagtaac 161

Qy 121 ttgaaacgtggacaaagccgccccacaccttcagaaacactcgttcacatacaaatc 180
Db 162 ttgaaacgtggacaaagccgccccacaccttcagaaacactcgttcacatacaaatc 221

Qy 181 cagatcgtgagatgacacccagggagcatttgaagtcttcacaaagactccttggaaggtg 240
Db 222 cagatcgtgagatgacacccagggagcatttgaagtcttcacaaagactccttggaaggtg 281

Qy 241 ctgctgagtcagtttgaacacacactcaactcgtggaagtgactatttggcctcgag 300
Db 282 ctgctgagtcagtttgaacacacactcaactcgtggaagtgactatttggcctcgag 341

Qy 301 ttctctgatcacaaaagatcacggttgctggtgatctctctaaacccatttgaaacag 360
Db 342 ttctctgatcacaaaagatcacggttgctggtgatctctctaaacccatttgaaacag 401

Qy 361 attagaaggccaaagacgctgtgttgaattgtggtgaaattcttccgcctgaccac 420
Db 402 attagaaggccaaagacgctgtgttgaattgtggtgaaattcttccgcctgaccac 461

Qy 421 acacaactccaagaaactcacaggtacactgtctcgtcgtcaggtgaagcaggacttg 480
Db 462 acacaactccaagaaactcacaggtacactgtctcgtcgtcaggtgaagcaggacttg 521

Qy 481 gctcaaggcaggttgacgtgtaaaacacacagcgcagctccttgatttcacacattgtg 540
Db 522 gctcaaggcaggttgacgtgtaaaacacacagcgcagctccttgatttcacacattgtg 581

Qy 541 caatctgagattgggatttttgatgaagcctttggacagagacacttagcaaaaaataaa 600
Db 582 caatctgagattgggatttttgatgaagcctttggacagagacacttagcaaaaaataaa 641

Qy 601 tacatactcagaagaacgactagagacaaaatcgtgaaatttcacataaccacatt 660
Db 642 tacatactcagaagaacgactagagacaaaatcgtgaaatttcacataaccacatt 701

Qy 661 ggacaaacacagcagatacagatttccagctcctagagattgcccgtcggttagagatg 720
Db 702 ggacaaacacagcagatacagatttccagctcctagagattgcccgtcggttagagatg 761

Qy 721 tatggaatccggttgaccccgccaaagacagggaaagcagagatcattctgcccgtt 780

```

```

Db 762 tatggaatccggttgaccccgccaaagacagaggagccagaaatcaatctggccgtt 821
Qy 781 gcaaacacgggaattctagttttcagggtttcactaaagatcaatgcttcaactggccc 840
Db 822 gcaaacacgggaattctagttttcagggtttcactaaagatcaatgcttcaactggccc 881
Qy 841 aagtgccggaagctgagcttcaagagaaagcgttttctcatcaagctccggccagatgcc 900
Db 882 aagtgccggaagctgagcttcaagagaaagcgttttctcatcaagctccggccagatgcc 941
Qy 901 aatagtcgtaccaggaataccttgaattcctgatgcccagtcgggatttctgcaatgcc 960
Db 942 aatagtcgtaccaggaataccttgaattcctgatgcccagtcgggatttctgcaatgcc 1001
Qy 961 tcttgaaaaatcgtgttgaacatcattccttcttagactttttgaagagcccaaacca 1020
Db 1002 tcttgaaaaatcgtgttgaacatcattccttcttagactttttgaagagcccaaacca 1061
Qy 1021 aagcccaagccctcctcttttagccgggggtcatcatttctggttcagttcggactcag 1080
Db 1062 aagcccaagccctcctcttttagccgggggtcatcatttctggttcagttcggactcag 1121
Qy 1081 aagcaggttctcgaactatgttaaagaaggaggacataaaggtgcagtttgaaaggaag 1140
Db 1122 aagcaggttctcgaactatgttaaagaaggaggacataaaggtgcagtttgaaaggaag 1181
Qy 1141 cacagcaagattcattctatccgagacgttgccttcacagctacagaaactgaattcggaa 1200
Db 1182 cacagcaagattcattctatccgagacgttgccttcacagctacagaaactgaattcggaa 1241
Qy 1201 gtgctgagcagtcctcagcagagcaccagccttacatttggagaaggtgcgcaactctcca 1260
Db 1242 gtgctgagcagtcctcagcagagcaccagccttacatttggagaaggtgcgcaactctcca 1301
Qy 1261 gggggccagagctgcgcggcaggaagaaagccttcgcgcggggagcgggggtcg 1320
Db 1302 gggggccagagctgcgcggcaggaagaaagccttcgcgcggggagcgggggtcg 1361
Qy 1321 caccagagcctgcgcggcaggaagaaagccttcgcgcggggagcggggagcgcgc 1380
Db 1362 caccagagcctgcgcggcaggaagaaagccttcgcgcggggagcggggagcgcgc 1421
Qy 1381 tcggcgccacgaggaaggaaggaaggttcgttaaggataggacccagcagagataaacct 1440
Db 1422 tcggcgccacgaggaaggaaggaaggttcgttaaggataggacccagcagagataaacct 1481
Qy 1441 cagcccccagcagcacaagcagcctcctcactgactggcagctcacttccgagctgtct 1500
Db 1482 cagcccccagcagcacaagcagcctcctcactgactggcagctcacttccgagctgtct 1541
Qy 1501 gtgaactcgcagggggagtgggccctgcacacgtgaccttctcccaacctgagcccc 1560
Db 1542 gtgaactcgcagggggagtgggccctgcacacgtgaccttctcccaacctgagcccc 1601
Qy 1561 gacaccaagcagcctcctccttgatcagccctgcctgctgtaatgaccagggcctgcccccg 1620
Db 1602 gacaccaagcagcctcctccttgatcagccctgcctgctgtaatgaccagggcctgcccccg 1661
Qy 1621 acggacgatgagatgagggccggaggaagagattcccaactgataaagcgtacttcata 1680
Db 1662 acggacgatgagatgagggccggaggaagagattcccaactgataaagcgtacttcata 1721
Qy 1681 gctaaagaaagtgtctaccacgcagcaacatactgagatctcgaagtattacacttcg 1740
Db 1722 gctaaagaaagtgtctaccacgcagcaacatactgagatctcgaagtattacacttcg 1781
Qy 1741 tggtttcagagcagtgagcaaaagagacgcatgcgcgaagcactgaaaaagtctcata 1800
Db 1782 tggtttcagagcagtgagcaaaagagacgcatgcgcgaagcactgaaaaagtctcata 1841
Qy 1801 ttcccgaattttgaacctttgcacaaatttcataactaatttctcgaaggaaattgagcaa 1860
Db 1842 ttcccgaattttgaacctttgcacaaatttcataactaatttctcgaaggaaattgagcaa 1901

```

QY 1861 cgaattgccctgtggaaggccgctcaaatgcccaaatcagaattaccaaagaatcggc 1920  
Db 1902 cgaattgccctgtggaaggccgctcaaatgcccaaatcagaattaccaaagaatcggc 1961  
QY 1921 gatgtcatgctgaagaacatcagggcgtgaagcacctggcggctcacctgtggaagcac 1980  
Db 1962 gatgtcatgctgaagaacatcagggcgtgaagcacctggcggctcacctgtggaagcac 2021  
QY 1981 agcaggccttgaggccctggagaatggaatcaagagctcccgcggctggagaacttc 2040  
Db 2022 agcaggccttgaggccctggagaatggaatcaagagctcccgcggctggagaacttc 2081  
QY 2041 tgcagagactttgagctgcagaagggtgtttacctacgcgtcaaacacttccctcgtggg 2100  
Db 2082 tgcagagactttgagctgcagaagggtgtttacctacgcgtcaaacacttccctcgtggg 2141  
QY 2101 ccaetgcaccggctcatgcactacaagcaggtcctcggagcgtgtgcaaacaccaccg 2160  
Db 2142 ccaetgcaccggctcatgcactacaagcaggtcctcggagcgtgtgcaaacaccaccg 2201  
QY 2161 ccgagccacgcgacttcagggactgcgagccgctttggcagagatcacggagatggtg 2220  
Db 2202 ccgagccacgcgacttcagggactgcgagccgctttggcagagatcacggagatggtg 2261  
QY 2221 gcaagctccacggttacgatgataagatgagaatttccagaagctgcagaaactcaag 2280  
Db 2262 gcaagctccacggttacgatgataagatgagaatttccagaagctgcagaaactcaag 2321  
QY 2281 aaagatttgattggcatgacaactctgtgttcgggaaaggagttcatcgtctgggc 2340  
Db 2322 aaagatttgattggcatgacaactctgtgttcgggaaaggagttcatcgtctgggc 2381  
QY 2341 agcctcagaagctctcgggaagggtccagcagcgtatgttctcgtttcaacgac 2400  
Db 2382 agcctcagaagctctcgggaagggtccagcagcgtatgttctcgtttcaacgac 2441  
QY 2401 gtcctgtatacagcagccggggtgcgcgcctccaatcagtttaaaagtccacggcag 2460  
Db 2442 gtcctgtatacagcagccggggtgcgcgcctccaatcagtttaaaagtccacggcag 2501  
QY 2461 ctccgcgtctatggcatgacgatgagagagcagagcagtggtggggtgcccactgc 2520  
Db 2502 ctccgcgtctatggcatgacgatgagagagcagagcagtggtggggtgcccactgc 2561  
QY 2521 ctgacctccggggccagcgcagtcctcatcgttgcgcgcagttctcgttccgagatg 2580  
Db 2562 ctgacctccggggccagcgcagtcctcatcgttgcgcgcagttctcgttccgagatg 2621  
QY 2581 gagaagtgggttggagacatccagatggcattgacctggcgggagaagcagcagcccc 2640  
Db 2622 gagaagtgggttggagacatccagatggcattgacctggcgggagaagcagcagcccc 2681  
QY 2641 gcccttgatttcttgccagcagccccctgacaaagttcccttgatgaagccaccgcg 2700  
Db 2682 gcccttgatttcttgccagcagccccctgacaaagttcccttgatgaagccaccgcg 2741  
QY 2701 gctgaccaggagtcagagatgacctgagcgcctcgcacatcgtgagcgcagggcc 2760  
Db 2742 gctgaccaggagtcagagatgacctgagcgcctcgcacatcgtgagcgcagggcc 2801  
QY 2761 ccgcaaccggcgaacacaaatggtgcacgtgtgctggcaccgcaaacaccagcgtctccatg 2820  
Db 2802 ccgcaaccggcgaacacaaatggtgcacgtgtgctggcaccgcaaacaccagcgtctccatg 2861  
QY 2821 gtggacttcagcatcgcagtgagaatcagttgtgtgaaacctgtgagaaattcaaa 2880  
Db 2862 gtggacttcagcatcgcagtgagaatcagttgtgtgaaacctgtgagaaattcaaa 2921  
QY 2881 aacagcaacgggtggcagaagctgtgggtgtgtgttcacaaacttctcgtcttcttac 2940  
Db 2922 aacagcaacgggtggcagaagctgtgggtgtgtgttcacaaacttctcgtcttcttac 2981

QY 2941 aaatcacaccaggacaatcatccttgcagcgtcgtctgtcgtcgtcgtcgtcgtcgtc 3000  
Db 2982 aaatcacaccaggacaatcatccttgcagcgtcgtctgtcgtcgtcgtcgtcgtcgtc 3041  
QY 3001 atccctctgagtcgagaaacatccagaaagactacgtgttcaagctgcacctcaagtc 3060  
Db 3042 atccctctgagtcgagaaacatccagaaagactacgtgttcaagctgcacctcaagtc 3101  
QY 3061 cacgtctactcttcagggcggaagcaggtacacgtttcgaaaagtgatggaagtgc 3120  
Db 3102 cacgtctactcttcagggcggaagcaggtacacgtttcgaaaagtgatggaagtgc 3161  
QY 3121 cgcagtcgaccagctctgcctcgcgacccccacgtgttgagccacaagagctctctgtg 3180  
Db 3162 cgcagtcgaccagctctgcctcgcgacccccacgtgttgagtcacaagagctctctgtg 3221  
QY 3181 tattgatggcggaacacactcgtttcccgagtggtgttctcgtggaagacgtttccctt 3240  
Db 3222 tattgatggcggaacacactcgtttcccgagtggtgttctcgtggaagacgtttccctt 3281  
QY 3241 ttctgttattaatgaagcctgtgtaaaattaacacactgtctgaaaatcaaaaaacatggctt 3300  
Db 3282 ttctgttattaatgaagcctgtgtaaaattaacacactgtctgaaaatcaaaaaacatggctt 3341  
QY 3301 ccagcagctctcctgttccacagccgggttttttaacccccgcacctctcagcgtttgaa 3360  
Db 3342 ccagcagctctcctgttccacagccgggttttttaacccccgcacctctcagcgtttgaa 3401  
QY 3361 tgaacagcgtcccccacactccagctcgtcgtcgtggggtggttcttcttagctagt 3420  
Db 3402 tgaacagcgtcccccacactccagctcgtcgtcgtggggtggttcttcttagctagt 3461  
QY 3421 ccagttataaaacattgtcatt 3442  
Db 3462 ccagttataaaacattgtcatt 3483

RESULT 3  
AAC98992  
ID AAC98992 standard; cDNA; 3094 BP.  
XX  
AC AAC98992;  
XX  
DT 09-MAR-2001 (first entry)  
XX  
DE Human pancreatic cancer antigen nucleotide sequence SEQ ID NO:220.  
XX  
KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;  
KW detection; diagnosis; identification; cytostatic; neuroprotective;  
KW neotropic; immunomodulatory; relaxant; contraceptive; gynaecological;  
KW antiinflammatory; cardiant; gene therapy; chromosome mapping;  
KW linkage analysis; tissue identification; tissue typing; forensic;  
KW neural; immune system; muscular; reproductive; gastrointestinal;  
KW pulmonary; cardiovascular; renal; proliferative; ss.  
OS Homo sapiens.  
XX  
FN WO200055320-A1.  
XX  
PD 21-SEP-2000.  
XX  
PF 08-MAR-2000; 2000WO-US05989.  
XX  
PR 12-MAR-1999; 99US-0124270.  
XX  
PA (HUMA-) HUMAN GENOME SCI INC.  
XX  
PI Rosen CA, Ruben SM;  
XX  
DR WPI; 2000-579444/54.  
XX  
P-PSDB; AAB54227.  
PT New nucleic acid that is a pancreatic cancer antigen for preventing,



PT treating, or ameliorating a medical condition, particular pancreatic  
PT cancer, or for use in assays for diagnosing a pathological condition -  
XX  
PS Claim 1; Page 664-665; 1379pp; English.

AAC98773 to AAC99231 encode the human pancreatic cancer associated proteins, called pancreatic cancer antigens, given in AAB54008 to AAB54466. The human pancreatic cancer antigens have cytostatic, neuroprotective, nootropic, immunomodulatory, relaxant, contraceptive, gynaecological, cardiant and antiinflammatory activities, and can be used in gene therapy. The polynucleotide and proteins can be used for preventing, treating, or ameliorating a medical condition or in assays for diagnosing a pathological condition or a susceptibility to one in a subject. Binding partners to the proteins and the activity of the proteins can be identified. The pancreatic cancer antigens can be used to detect, treat or prevent pancreatic disorders, especially cancer. Agonists and antagonists to the antigens can be screened for. The pancreatic cancer antigen polynucleotides can be used to design nucleic acid hybridisation probes that can be used in chromosome mapping, linkage analysis, tissue identification and/or typing and a variety of forensic and diagnostic methods. The proteins can be used to generate antibodies which are used to purify, detect and target the polypeptides, including both in vivo and in vitro diagnostic and therapeutic methods. The proteins can be used to treat or prevent neural, immune system, muscular, reproductive, gastrointestinal, pulmonary, cardiovascular, renal or proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent sequences used in the exemplification of the present invention.

Sequence 3094 BP; 736 A; 849 C; 755 G; 749 T; 5 other:

Query Match 49.1%; Score 1688.6; DB 21; Length 3094;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1699: Conservative 2; Mismatches 1; Indels 1; Gaps 1;

QY	1740	gtggttttcagagcacagttagcaagagagacgccatgcggaagcaactgaaagtctcat	1799
DB	62	gtggttttcagagcacagttagcaagagagacgccatgcggaagcaactgaaagtctcat	121
QY	1800	attcccgaaatttgaaccttgcacaaatttcataactaatttctcaaggaaatttgagca	1859
DB	122	attcccgaaatttgaaccttgcacaaatttcataactaatttctcaaggaaatttgagca	181
QY	1860	acgacttgccctgtggaagggccgtccaattgcccaatcagagattaccaaaaaatctgg	1919
DB	182	acgacttgccctgtggaagggccgtccaattgcccaatcagagattaccaaaaaatctgg	241
QY	1920	cgatgtcatgctgaagaacattcagggcataaagcacactgcgcgtcaccttgtggaagca	1979
DB	242	cgatgtcatgctgaagaacattcagggcataaagcacactgcgcgtcaccttgtggaagca	301
QY	1980	cagcagagcccttgagggccctggaaatggaaatcaagagctcccgcgctggagaaactt	2039
DB	302	cagcagagcccttgagggccctggaaatggaaatcaagagctcccgcgctggagaaactt	361
QY	2040	ctgcagagactttgagactcagaaggtgtgtttacctacccgtctaacaccttctctctcg	2099
DB	362	ctgcagagactttgagctcagaaggtgtgtttacctacccgtctaacaccttctctctcg	421
QY	2100	ggcaactgcaccggctcatcactacaagcaggtcctggagcggctgtgcaaacaccaccc	2159
DB	422	ggcaactgcaccggctcatcactacaagcagggtctctggagcggctgtgcaaacaccaccc	481
QY	2160	gccgagccacgcgacttcagggaactgcgcagcgcctttggcagagatcacggagatgtgt	2219
DB	482	gccgagccacgcgacttcagggaactgcgcagcgcctttggcagagatcacggagatgtgt	541
QY	2220	ggcacagcttccacgttacatgatcaagatgagaatttcagaagctgcacgaactcaa	2279
DB	542	ggcacagcttccacggcacatgatcaagatggagaatttcccagaagctgcacgaactcaa	601
QY	2280	gaaagatttgattggcattgacaactctgtggttcggggaaggagttcataccgctctggg	2339

Db	602	gaaagatttgattggaattgacaactcttggtgtcccggaaggaggttcaatcgtctctggg	661
Qy	2340	cagctcagcaagctctcggggaagggtccacgagcgcattgtcttcctgttcaacga	2399
Db	662	cagctcagcaagctctcggggaagggtccacgagcgcattgtcttcctgttcaacga	721
Qy	2400	cgtcctgctatacaagcagccgggggtgacgagcctccaatcagtttaaagtccacggga	2459
Db	722	cgtcctgctatacaagcagccgggggtgacgagcctccaatcagtttaaagtccacggga	781
Qy	2460	gctcccctctatggatcagctatcagagagcgaagacagatgggggggtgcccactg	2519
Db	782	gctcccctctataggcatgacgattgagagagcgaagacagatgggggggtgcccactg	841
Qy	2520	cctgacccctcggggccagcgagtcacatcctgtgtggccgcagcttctcggtcgagat	2579
Db	842	cctgacccctcggggccagcgagtcacatcctgtgtggccgcagcttctcggtcgagat	901
Qy	2580	ggagaatgggttgagacatccagatggccattgacctggcggagaagcagcagccc	2639
Db	902	ggagaatgggttgagacatccagatggccattgacctggcggagaagcagcagccc	961
Qy	2640	cgccctcagttcctggccagcagcccccctgcacaagaatccccgtatgaagccaacgc	2699
Db	962	cgccctcagttcctggccagcagcccccctgcacaagaatccccgtatgaagccaacgc	1021
Qy	2700	ggctgcacagagtcagagatgacctgagccctcgccacatcgctggagcgcagcc	2759
Db	1022	ggctgcacagagtcagagatgacctgagccctcgccacatcgctggagcgcagcc	1080
Qy	2760	ccgcacccgcgcaacacaaatggtgcacgtgtgtggcaccgcaacacagcgtctccat	2819
Db	1081	ccgcacccgcgcaacacaaatggtgcacgtgtgtggcaccgcaacacarctctccat	1140
Qy	2820	ggtggaattcagcatcgcagtgagaaatcagttgttggaaacctgctgaggaattcaa	2879
Db	1141	ggtggaattcagcatcgcagtgagaaatcagttgttggaaacctgctgaggaattcaa	1200
Qy	2880	aaacagcaacgggtgcagaagctgtggtgtgttccaaaacttcgctgtctctcta	2939
Db	1201	aaacagcaacgggtgcagaagctgtggtgtgttccaaaacttcgctgtctctcta	1260
Qy	2940	caaatcacacaggaacaaatcccccctgcagcctgcctctgctcggtactcgtctcac	2999
Db	1261	caaatcacacaggaacaaatcccccctgcagcctgcctctgctcggtactcgtctcac	1320
Qy	3000	catcccctctgagtcgcagaaacatccagaaaactactgttccaagctgcactccaagtc	3059
Db	1321	catcccctctgagtcgcagaaacatccagaaaactactgttccaagctgcactccaagtc	1380
Qy	3060	ccagctctactactcagggcgaaagcagatcacggttcgaaaggtgagatggaagtgat	3119
Db	1381	ccagctctactactcagggcgaaagcagatcacggttcgaaaggtgagatggaagtgat	1440
Qy	3120	ccgcagtgccaccagctctgcttcgcgcgacccacagctgttgagccacaagagctcttgt	3179
Db	1441	ccgcagtgccaccagctctgcttcgcgcgacccacagctgttgagccacaagagctcttgt	1500
Qy	3180	gtattgatggccggacacactggtttccgcagtggtcttctcctggaagacgtttccct	3239
Db	1501	gtattgatggccggacacactggtttccgcagtggtcttctcctggaagacgtttccct	1560
Qy	3240	tcttctgtataatgaagcctggtaaaaataacacctgtctgaaaaatcaaaaacatggct	3299
Db	1561	tcttctgtataatgaagcctggtaaaaataacacctgtctgaaaaatcaaaaacatggct	1620
Qy	3300	tcccagcagctctcctgtctccacagcccggttttttaaccccgcacctctcagcgtttga	3359
Db	1621	tcccagcagctctcctgtctccacagcccggttttttaaccccgcacctctcagcgtttga	1680
Qy	3360	atgaacagcgtccccacctccagtccttgcatccgcctggggggcgctgttctttagctagt	3419
Db	1681	atgaacagcgtccccacctccagtccttgcatccgcctggggggcgctgttctttagctagt	1740



CC (Alzheimer's, Parkinson's, Huntington's disease, stroke), coagulation

XX

AAA0581-96 encode human cytoskeleton associated proteins 1 to 16  
CC (CYSKP-1 to CYSKP-16) respectively. The sequences can be used to treat  
CC and diagnose cancer and cell proliferative, autoimmune/inflammatory,  
CC vesicle trafficking, neurological, cardiovascular, cell motility,  
CC reproductive and muscle disorders. Pharmaceutical compositions  
CC containing CYSKP-2 to CYSKP-16 can be used to treat or prevent disorders  
CC associated with decreased expression or activity of CYSKP (claimed), for  
CC example, atherosclerosis, cirrhosis, hepatitis, myelofibrosis, psoriasis,  
CC cancers, autoimmune/antiinflammatory disorders such as allergies, anemia,  
CC asthma, acquired immunodeficiency syndrome (AIDS), Crohn's disease,  
CC diabetes mellitus, Goodpasture's syndrome, osteoarthritis, scleroderma  
CC and trauma. CYSKP antagonists can be used to treat or prevent a  
CC disorder associated with increased expression or activity of CYSKP  
CC (claimed).  
xx  
xx sequence 3620 BP; 1156 A; 773 C; 906 G; 785 T; 0 other;

Query Match 4.9%; Score 169.6; DB 21; Length 3620;  
Best Local Similarity 51.4%; Pred. No. 1.5e-35;  
Matches 472; Conservative 0; Mismatches 434; Indels 12; Gaps 3;

Qy	175	aaatccagatgctggatgacacccaggaggcaatttgaagttccacaagaagctcctggg	234
Db	729	aaagtacccctcttagatggcaccgaatacagcgtgacctggagaacaatgccaaggga	788
Qy	235	aaagtgcctggatgcagtttgcacacacctaactcgaggaggtgactattttgc	294
Db	789	caagtgtatttgacaagtgtggaacacctaactctctggagaagagctactttgga	848
Qy	295	ctcgagtttctctgatcacaaaagatcacggtgtggctggatctctctaaaccattgtg	354
Db	849	cttttgttcaggaaagccctgagcagaaaactgttagatcctgctaaagaaaataag	908
Qy	355	aaacagattagaagggccaaagcacgttgttgaagtttggtgaaattctttccgct	414
Db	909	agacaactgagaacc---ttccatggctattacttttaatgtgaagttttactctct	965
Qy	415	gaccacacacactccaagaagaactcacaaggtacctgttcgcgctcaggtgaagcag	474
Db	966	gatcctctcaattgactgaagatatcacagatactctctgtgcttcagctccggcag	1025
Qy	475	gacttggctcaagcaggttgacgtgtaatgacaccagcgagctcttctgatttcacac	534
Db	1026	gacattgcctctggcgcctgcctgtctttgtgactcatgtctctctggatccctac	1085
Qy	535	atttgcgaactgagattggggatttt-----gatgaagccttggacagagacactta	588
Db	1086	acctgcaggtgaacttggtagctatgaccgaagaacatggcagcatcgacctcagt	1145
Qy	589	gcaaaaataaatcacatactctcgaagacgcactagaggacaaaatctgtgaatttcac	648
Db	1146	gaattccagtttgcctctactcagactaaggagctggaaagaaaggtggcagagctcac	1205
Qy	649	cataaccacattggacaaaaccacgagaaatcagatttccagctcctagagatgcccgt	708
Db	1206	aaacccacagggttatctgcgcagcacaagtgtatcccaagtcttagaaaaatgcagaag	1265
Qy	709	cggtagagatgtatgaatccggttgaccgccggccaaaggacagggaaagccagaagatc	768
Db	1266	aggtttccatgtatgtgtgaectacatcatgccaaaggactcagaaggtgtggacatc	1325
Qy	769	aatctggccgttgccaacacggaattctagtgtttcagggtttcactaaatcaatgcc	828
Db	1326	aagctggggcgtgtgtgtatggactctctattacaaaagacagactgcgaatcaatcgt	1385
Qy	829	ttcaactgggcccagggtgcggaagctgagcttcgaaggaagcgtttctctcatcaagctc	888
Db	1386	tttgcctggccgaaaaacttaaaaatttccctataaaacgcagtaacttctacataaagtc	1445
Qy	889	cgccagatgccaatagtgtgtacaggatataccttggaaattcctgtatggccagtcgggat	948
Db	1446	agaccggcagagcttggaacagtttgadagttaccatttgattcaactctccaaaccacqg	1505

Qy	949	tcttcaagtcctcttggaatactcgtgttgaaacatcatgcctttcttagaactttgaa	1008
Db	1506	gcagcgaaaagaactatggaaaagtgcgtggagcatcatactttctcacaggcttgt---	1562
Qy	1009	gagcccaaaccaaaagcccagccgtcctctcttttagccgggggtcatcattcgggttcagt	1068
Db	1563	tctccagcagcagccaccaaaagccaagttctctgacctggggtgccaaatttcgatagt	1622
Qy	1069	ggtcggactcagaagcag; 1086	
Db	1623	ggccgcaccacaagcag 1640	
RESULT	6		
AAH98023			
ID	AAH98023 standard; DNA; 544 BP.		
XX			
AC	AAH98023;		
XX			
DT	10-OCT-2001 (first entry)		
XX			
DE	Murine 7-transmembrane G-protein coupled receptor coding sequence #267.		
XX			
KW	Murine; stromal stem cell; signalling; vaccine; 7TM-GPCR;		
KW	7-transmembrane G-protein coupled protein receptor; ds.		
OS	Mus sp.		
PN	WO200160999-A1.		
XX			
PD	23-AUG-2001.		
XX			
PF	14-FEB-2001; 2001WO-US04700.		
XX			
PR	14-FEB-2000; 2000US-0182377.		
XX	(IMCL-) IMCLONE SYSTEMS INC.		
PA	(UYPR-) UNIV PRINCETON.		
PI	Lemischka IR, Witte L, Pereira DS;		
DR	WPI; 2001-522596/57.		
PT	DNA Sequences encoding 7-transmembrane G-protein coupled protein		
PT	receptors characteristic of hematopoietic stem cells, useful for		
PS	treating leukemia -		
XX	Claim 1; Page 122; 176pp; English.		
CC	The present invention relates to murine coding sequences for		
CC	7-transmembrane G-protein coupled protein receptors (7TM-GPCRs). The		
CC	present sequence is one such murine 7TM-GPCR coding sequence. The present		
CC	sequence was derived from stromal stem cells. The present sequence		
CC	and its corresponding protein are useful in the prevention, diagnosis and		
CC	treatment of diseases associated with inappropriate 7TM-GPCR expression.		
CC	7TM-GPCRs identify specific signalling molecules, to activate an		
CC	effector-signalling cascade that triggers an intracellular response and		
XX	eventually a biological effect.		
S	Sequence 544 BP; 143 A; 131 C; 150 G; 120 T; 0 other:		

Query Match	4.4%;	Score 151.6;	DB 22;	Length 544;
Best Local Similarity	56.3%;	Pred. No. 4.2e-31;		
Matches 304;	Conservative	0;	Mismatches 234;	Indels 2; Gaps 1;

  

y	2279	agaaagatttgatggcattgacaatctctggttcggggaaggaggtccatccgtctgg	2338
b			
	4	agcgagacctggtcgggtgtagaaaacctcatgctctctgggaggaggttaccgtatgagg	63
y	2339	gcagccttcagcaagctctcggggaaggggctccacgacgcagctgttcttcctgttcaacg	2398





allogenic transplant rejection; multiple myeloma; Burkitt's lymphoma; leukaemia; cancer; neoplasms.

Homo sapiens.

Key	Location/Qualifiers
CDS	230..1339

```

FTI
FT /lag= d
FT /product= "Human PTPase #2"

```

WO200153530-A1.

26-JUL-2001.

17-JAN-2001:

18-JAN-2000: 2000US-0176306

XX  
PA  
(HUMA-) HITMAN GENOME SCT INC.

XX  
PT chf v Dubon SM.

XX  
DR WPT: 2001-442265/

DR P-PSDB; AAU04486.  
YY

PT New human protein

0  
8  
6  
  
7  
1  
7  
3

XX  
XX

XX 1

CC (PTPase) polypept

conditions are di

CC (c.g. Alzheimer's  
CC Tourette's syndrome)

CC  
CC  
CC  
CC

CC sclerosis, ankylo-

CC vitiligo, asthma,

disorders (e.g. p  
rejection) and ca

CC and multiple myel  
nucleic acid enco

CC susceptibility to  
CC in screening for

XX  
50  
sequence 2872 bp.

Query Match

**Matches 490; Conse**

Qy 131 gacagaagccgc

Db 1 ggcacgagccgc

Qy 191 atgacaccagg

Db 60 acgggaccgaag

Qy 251 cagtttgcaacc

Db 120 agattgtgtacc

311 acaaaaqatca Qy

180 ctaccaggtta

371 caaacacattc

[illegible]

	Qy	431	aagaagaaactcaacaaggtaacctgttccgcgctgcaggtgaagcaggagacttggtccaaggca	490
	Db	300	gtgagggagttaaacaagttacactgttgtttttaacaactcaggcatgacattcttcttcggaa	359
	Qy	491	ggttgacggttaataagacaaccagcgagcagctctcttgattcacacattgtgcaatctcgaa	550
	Db	360	aattgaaatgcccttaTgaacagctgtggaattagctctctctgtctacacgcggagc	419
	Qy	551	ttagggattttagaagccttggaaca--gagagcaacttagcaaaaaataaatatcatac	607
	Db	420	ttagggagtgcgagcttccagaacacacacagagccttgctgtagtttcgggttcattc	479
	Qy	608	c---tcagcaagacgcctagaggaacaaaatcgtggaaatttcaccataaacacattggac	664
	Db	480	caaatacagacaagaacaatggaatttgatatcttccagagatggaagaagtgcaggggaa	539
	Qy	665	aaacacacagagaatacagattccagctccttagagattgccctcgcggtagagatgatg	724
	Db	540	agagccctgccagcgcggaactctcctatctgtaataaagcgaagtggcttggaatgatg	599
	Qy	725	gaatccggttgcccccgccaaagacaggaagcagcaagaatcaatctggccgttgcca	784
	Db	600	gggtagacatgcacgttgctcaggggaagagatggctgTgaatatctcttggactgacc	659
	Qy	785	acacgggaattctagtgtttcagggtttcctaagatcaatgccttcaactgggccaaag	844
	Db	660	cgacaggcataattaacttttgaaggagcTaacaaaataggcttatctcttggcctaaaa	719
	Qy	845	tgcggaaagctgagcttccaagaggaagcgcttctctatcaagctccgcgcagatgccaa	904
	Db	720	ttaccaaaatggattttaaaagagcaaatgcacctcgtgggtgcgagatgatgc	779
	Qy	905	-----gtgcgtaccagataccttggaattctctgatggccagctcggtattctgcaag	958
	Db	780	agggacgtlgagcaagacacacgtttgTtccggttagacagtgcoaggaacctgcaac	839
	Qy	959	cctctctggaaaatactgtgtgaacatcatgcctctcttagacttttgaagagcccaac	101
	Db	840	acctttggaagtgtagtgagcaccacgacattcttcgactgcggagccaggaaca	899
	Qy	1019	caaaagcccac---gcccgctcctctttagccgggggttcattctcagttcagtggtcgga	107
	Db	900	gcaataccaatagatccgactttatcaggtgggctcgtctcagattcagttggcgga	959
	Qy	1076	ctcagaagcaggtt 1089	
	Db	960	cagaatatcaagct 973	
		RESULT 10.		
		AAH17787		
	ID	AAH17787 standard; cDNA; 3166 BP.		
	XX			
	XX	AAH17787;		
	XX			
	DT	26-JUN-2001 (first entry)		
	XX			
	DE	Human cDNA sequence SEQ ID NO:17434.		
	XX			
	KW	Human; primer; detection; diagnosis; antisense therapy; gene therapy;		
	OS	Homo sapiens.		
	XX			
	PN	EP1074617-A2.		
	XX			
	XX	07-FEB-2001.		
	XX			
	PF	28-JUL-2000; 2000EP-0116126.		
	XX			
	PR	29-JUL-1999; 99JP-0248036.		
	PR	27-AUG-1999; 99JP-0300253.		
	PR	11-JAN-2000; 2000JP-0118776.		

[illegible]







Db 207 agacaatttcagaaacttatgcaaatcagttacagcttaataatggacacattgaaattgt 266  
 Qy 2310 ggttcctggcgaagaggttcctcgtcgtggcagcctcagcaagctctcgggggaagggtc 2369  
 Db 267 gcagcctggcgggttttctcctcaagaagaattctgataagctgtctcgaaagtgtat 326  
 Qy 2370 ccagcagcgaatgttctcctgtttcaacagcgtcctctgtatatacagcggcggtgtgac 2429  
 Db 327 gcaactcgaagtgttctcctgttttaattgctcctgtgtatatacacaacacagtgacgtc 386  
 Qy 2430 ggcctccaatcagtttaaaagtcacagcggcagctcccgctctatgcatgacgattgagga 2489  
 Db 387 tggagat-----gtataaactgaacaacatgctctcactggtcggaatgaaggcagaaaa 440  
 Qy 2490 gagcgaagcagtggtgggggtgcccactgctgacccctcggggcagcggcagtcctcat 2549  
 Db 441 acctaccagaagcctatcagaatgaattaaagattgaaagttagaacg---ttcctt 497  
 Qy 2550 catcgtggcgcagctgtcgtccgagatggagaaagtggttgagacatccagatggc 2609  
 Db 498 cattctctcagccagttctgcccagaaaggatgaatggttagaagcgtattccagggc 557  
 Qy 2610 cattgacctggcggagaagcagcagcccc 2640  
 Db 558 aatagaagagtatgccaaagaaagaatcacc 588

RESULT 13  
 AAH15014  
 ID AAH15014 standard; cDNA; 1821 BP.  
 XX AC AAH15014;  
 XX DT 26-JUN-2001 (first entry)  
 XX DE Human cDNA sequence SEQ ID NO:12971.  
 XX KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX OS Homo sapiens.  
 XX PN EP1074617-A2.  
 XX PD 07-FEB-2001.  
 XX PF 28-JUL-2000; 2000EP-0116126.  
 XX PR 29-JUL-1999; 99JP-0248036.  
 XX PR 27-AUG-1999; 99JP-0300253.  
 XX PR 11-JAN-2000; 2000JP-0118776.  
 XX PR 02-MAY-2000; 2000JP-0183767.  
 XX PR 09-JUN-2000; 2000JP-0241899.  
 XX PA (HELI-) HELIX RES INST.  
 XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 XX PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX DR WPI; 2001-318749/34.

Primer sets for synthesizing polynucleotides, particularly the 5602 full-length cDNAs defined in the specification, and for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs -

Claim 8; SEQ ID 12971; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.

XX Sequence 1821 BP; 594 A; 344 C; 369 G; 514 T; 0 other;

Query Match 2.5%; Score 85; DB 22; Length 1821;  
 Best Local Similarity 49.4%; Pred. No. 1e-12;  
 Matches 282; Conservative 0; Mismatches 280; Indels 9; Gaps 2;

Qy 2070 ttactacgcctcaacacaccttctcctcgcggccactgcacgcgtctatgcactacaagca 2129  
 Db 27 taatctggcctcaagcactacctgctcaagcgggttcagaggatccccagtcacaggct 86

Qy 2130 ggtcctggagcgtgtgcaaacaccacccgcgcagccacgcagcttcaggagctcgcg 2189  
 Db 87 gttgctgacagattatttgaaatctcatagaagatgctggagattacagacacatca 146

Qy 2190 agcgcgtttgcagagatcacggagatggtggcacagctccacggtacgatgatacaat 2249  
 Db 147 agatgccttgctgtgttatagaggtagcccaaccacgccaatgacacacatgaagcaag 206

Qy 2250 ggagaatttccagaagctgcacgaactcaagaagattgttggtcattgacaaatttgt 2309  
 Db 207 agacaactttcagaataattatgcaaatcagtcacagcttaaatggacacatgaaattgt 266

Qy 2310 ggttcgggaagaggttccatccgtcgtggcagcctcagcaagctctcgggaagggtc 2369  
 Db 267 gcagcctggtcggtgtttctcaagaagaattctgtatgaagctgtctcggaaaagtgt 326

Qy 2370 ccagcagcgtatgttcttctgttcaacgacgtcctctctatatacagcggcggtgtgac 2429  
 Db 327 gcaactcgaatgttttctgtttaaagtatgctcgtgtatatacacaacacagtgacgtc 386

Qy 2430 ggcctccaatcagtttaaaagtcacagcggcgagctcccgctctatgcatgacgattgagga 2489  
 Db 387 tgggat-----gtataaactgaacaacatgctctcactggtggaatgaagtcagaaa 440

Qy 2490 gagcgaagcagtggtgggggtgcccactgcctgacccctcggggcagcggcagtcctat 2549  
 Db 441 acctaccagaagcctatcagaatgaattaaagattgaaagttagaacg---ttcctt 497

Qy 2550 catcgtggcgcaggttctcgtccgagatggagaagtggttgagacatccagatggc 2609  
 Db 498 cattctcagcaggttctgcccagaaaggatgaatggtgctagaagcgtattccagggc 557

Qy 2610 cattgacctggcggagaagcagcagcccc 2640  
 Db 558 aatagaagagtatgccaaagaaagaatcacc 588

RESULT 14  
 AAC69420  
 ID AAC69420 standard; cDNA; 410 BP.  
 XX AC AAC69420;  
 XX DT 30-JAN-2001 (first entry)  
 XX



DE Human secreted protein gene 22 SEQ ID NO:32.

XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;

KW antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;

KW cerebroprotective; nontropic; neuroprotective; antibacterial; virucide;

KW fungicide; ophthalmological; gene therapy; autoimmune disease; infection;

KW hyperproliferative disorder; cardiovascular disorder; angiogenesis;

KW cerebrovascular disorder; nervous system disorder; ocular disorder;

KW wound healing; skin aging; food additive; preservative; ss.

XX

OS Homo sapiens.

XX WO200058468-A2.

PN

XX 05-OCT-2000.

PD

XX 22-MAR-2000; 2000WO-US07526.

PF

XX 26-MAR-1999; 99US-0126600.

PR

XX 22-DEC-1999; 99US-0171550.

PR

XX (HUMA-) HUMAN GENOME SCI INC.

PA

XX Rosen CA, Ruben SM, Komatsoulis G;

PI

XX WPI; 2000-611713/58.

XX P-PSDB; AAB38140.

DR

XX Nucleic acids encoding human secreted proteins, used to prevent, treat,

PT ameliorate, or diagnose conditions such as autoimmune disorders, skin

PT disorders and cancer -

XX

PS Claim 1; Page 328; 374pp; English.

XX

CC The polynucleotide sequences given in AAC69399 to AAC69445 encode the

CC human secreted proteins given in AAB38119 to AAB38165. AAB38166 to

CC AAB38201 represent human secreted polypeptide sequences and proteins

CC homologous to them, which are given in the exemplification of the present

CC invention. Human secreted proteins have activities based on the tissues

CC and cells the genes are expressed in. Example of activities include:

CC immunosuppressive; antirheumatic; antirheumatic; antiproliferative;

CC cytostatic; cardiant; vasotropic; cerebroprotective; nontropic;

CC neuroprotective; antibacterial; virucide; fungicide; and

CC ophthalmological. The polynucleotides and polypeptides can be are used to

CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,

CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used

CC in diagnosing a pathological condition or susceptibility to a

CC pathological condition. Disorders which are diagnosed or treated include

CC autoimmune diseases, hyperproliferative disorders, cardiovascular

CC disorders, cerebrovascular disorders, angiogenesis, nervous system

CC disorders, infections caused by bacteria, viruses and fungi and ocular

CC disorders. The polypeptides can also be used to aid wound healing and

CC epithelial cell proliferation, to prevent skin aging due to sunburn, to

CC maintain organs before transplantation, for supporting cell culture of

CC primary tissues, to regenerate tissues and in chemotaxis. The

CC polypeptides can also be used as a food additive or preservative to

CC increase or decrease storage capabilities. AAC69390 to AAC69398 and

CC AAB38118 represent sequences used in the exemplification of the present

CC invention.

XX

SQ Sequence 410 BP; 100 A; 93 C; 105 G; 112 T; 0 other;

Query Match 2.2%; Score 76; DB 21; Length 410;

Best Local Similarity 94.0%; Pred. No. 1.3e-10;

Matches 79; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1061 ggttcagtggtcgactcagaagcagggttcctcgactatgttaagaaggagcataaga 1120

|||||

Db 53 gtttcagtggtcgactcagaagcagggttcctcgactatgttaagaaggagcataaga 112

|||||

QY 1121 aggttcagtttgaagaagcaca 1144

|||||

Db 113 aggtgcagtttgaagaagtaagaga 136

RESULT 15

AAH99833

ID AAH99833 standard; cDNA; 1156 BP.

XX

AC AAH99833;

XX

DT 16-OCT-2001 (first entry)

XX

DE Human protein encoding cDNA sequence SEQ ID NO:568.

XX

KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;

KW antinflammatory; antirheumatic; antiarthritic; immunosuppressive;

KW antibacterial; endocrine; cardiant; central nervous system; virucide;

KW anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia;

KW antiaggregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;

KW dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic;

KW neuroprotective; antidepressant; nontropic; antiparkinsonian; infection;

KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;

KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;

KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;

KW genetic disease; haematopoietic disorder; platelet disorder; asthma;

KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;

KW allergic rhinitis; diabetes; multiple sclerosis; depression;

KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;

KW neurological disorder; ss.

XX

OS Homo sapiens.

XX

PN WO200153455-A2.

XX

PD 26-JUL-2001.

XX

PF 22-DEC-2000; 2000WO-US35017.

XX

PR 23-DEC-1999; 99US-0471275.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

XX

XX (HYSE-) HYSEQ INC.

PA

XX Tang YT, Liu C, Drmanac RT;

PI

XX WPI; 2001-457603/49.

DR

XX P-PSDB; AAM25892.

XX

PT Isolated human polynucleotides encoding polypeptides, useful for the

PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX

PS Claim 1; Page 674; 1217pp; English.

XX

CC AAH99166 to AAH99904 encode the human proteins given in AAM25225 to

CC AAM25963. The proteins can have activities based on the tissues and

CC cells they are expressed in, such as: antinflammatory; antirheumatic;

CC antirheumatic; immunosuppressive; antibacterial; endocrine; cardiant;

CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;

CC cardiovascular; antianaemic; antiaggregant; haemostatic; vulnery;

CC antiulcer; osteopathic; dermatological; antiallergic; antiasthmatic;

CC antidiabetic; cytostatic; neuroprotective; antidepressant; nontropic;

CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides

CC encoding them can be used in gene therapy, antisense therapy and vaccine

CC production. The proteins and polynucleotides are useful for screening for

CC agonists or antagonists of a protein and for the treatment and diagnosis

CC of disorders associated with the activity of a protein e.g. inflammation,

CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,

CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal

CC infections, autoimmunity, genetic diseases, haematopoietic disorders,

CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,

CC osteoporosis, severe combined immunodeficiency, eczema, allergic

CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,

CC Alzheimer's disease, Parkinson's disease, neurodegenerative and

CC neurological disorders.

XX

SQ Sequence 1156 BP; 381 A; 202 C; 255 G; 318 T; 0 other;

Query Match 2.0%; Score 69.4; DB 22; Length 1156;  
Best Local Similarity 50.1%; Pred. No. 1.4e-08;  
Matches 262; Conservative 0; Mismatches 246; Indels 15; Gaps 3;

Qy	656	acattgacaaacaccagcaaatcatttccagctccagcttagagattgccgctcggttag	715
Db	381	acagaggtcaaacaccgacagcagctgaaccattatctgaataaagcccaatggctag	440
Qy	716	agatgtatggaatccggtgtgcccgcgaagagcagcagcagcaaatctctg	775
Db	441	aaatgtatgggttgatgcattgtgtcaagctagagatgggaatgactatagtgtg	500
Qy	776	ccgttgcaacacgggaattctagtgtttcagggtttcactaagatcaatgcctcaact	835
Db	501	gactaacacacacagaggtcctgttttgaaggagatacccaaatggcttttttt	560
Qy	836	gggccaaaggtgcgaagctgaagcttcagaggaagcgtttctcatcaagctccggccag	895
Db	561	ggccgaagataaccagattggattttaagagaataaataaaccttggtgtgtagaag	620
Qy	896	atgccaat-----agtgcgtaccagatcaccttggaattcctgattgccagtcgggatt	949
Db	621	atgatgacagggaagaacagagcaacatacatcttcttagactgagatcatccaaaag	680
Qy	950	tctgaagtccttctgaaatctgtgtgaacatcatgctcttcttagactttttgaag	1009
Db	681	catgcaaacatttatggaatgtgtgtgagcatcatgtttcttcgcttcgagggcc	740
Qy	1010	agcccaaaccaagcccaagc---ccgtcctcttagccggggtcatcatttcggttca	1066
Db	741	ccgtcccaaaaggttctcatcgatcaggatttctcagtaggaggtacacgatttagata	800
Qy	1067	gtggtcggactcagaagcaggttctcagactatgttaaagaaggagagcataaagaggtgc	1126
Db	801	gtgggaaacagagtatcagaccacacaaacaaataaagcaagaagatcaa-----cat	854
Qy	1127	agtttgaaggaagacagcaagattctattctatccggagcct	1169
Db	855	cccttgaagaagggcccaacacgatatctctagacgaactct	897

Search completed: December 6, 2001, 12:00:11  
Job time: 11323 sec



Matches 218; Conservative 0; Mismatches 247; Indels 3; Gaps 1;

QY 178 atcagatgctgatgacacccaggagcatttgaagttccacaaagagctcctgggaag 237  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
Db 411 ATCAACTGCTTAAACAGAGTTGTGGAGTTACCCCTGCTGGAGACACTGGCCAG 470  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
QY 238 gtgctgctgatgcagtttgaacacccctcaacctcgtggaagtgactatttggcctc 297  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
Db 471 GAAAGCTCGAGCCGTCGCCAGAGCTGGAGCTGCGGAGGTCACTTACTTCAGCCTC 530  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
QY 298 gagtttctgatcacacaaagatacaogtggctgtggtctcctctaaacccattgtgaaa 357  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
Db 531 TGTACTACAAACAGCAAAATACAGCGCGGTGGGTAGATTGGAAAAACCTTTGAAGAAG 590  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
QY 358 cagat---tagaaggccaagcagctgttggtttaagtgttgtaaatcttccgcct 414  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
Db 591 CAGCTGATAAATATGATTTGGAACTTACCGTCTATTTTGGAGTGGTGTATGTGCGCT 650  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
QY 415 gaccacacacaaactccaagaagaactcacaaaggtacctgttcgctgcaggtgaagcag 474  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
Db 651 TCAGTTTCTCAGCTGCAGCAGAGATTTACCAGGTATCAGTATTATCTGCAACTGAAGAAA 710  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
QY 475 gacttggctcaagcaggttgtaogtgaatgacacacagcagctctcttcttgatttcacac 534  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
Db 711 GATATCTTGAAGGAAGATATTCCTGTACCTTAGAACAGCAATTCAGCTAGCAGGCTTA 770  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
QY 535 attgtgaactgatagattgggatttggatgaagccttggacagagagcacttagcaaaa 594  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
Db 771 GCTGTTCAAGCGGATTTTGTGTGATTTGATCAGTATGAATCCAGGACTTCTTTCAGAAA 830  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
QY 595 aataaatacatacctcagcagcagcactagagagacaaaatcgtggaa 642  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...  
Db 831 TTTGCCCTGTCTCTGTGGGATGTTTACAGATGAAAAGTATTGGAA 878  
|||||...|||||...|||||...|||||...|||||...|||||...|||||...|||||...

RESULT 2  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109

; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: PTZgpt-Fls  
; US-08-232-463-14

Query Match 1.6%; Score 53.8; DB 1; Length 7218;  
Best Local Similarity 4.7%; Pred. No. 0.00023;  
Matches 19; Conservative 222; Mismatches 164; Indels 0; Gaps 0;

QY 1058 ttcggttcadgtgcggaactcagaagcaggttctcgactatgttaaagaaggagacata 1117  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
Db 1437 TACRR 1378  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
QY 1118 agaagtgagtttgaagaagcacagcaagattctatccggagccttgcttcac 1177  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
Db 1377 RRR 1318  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
QY 1178 agctacagaaactcgaagtgctgagcagctcagcagcagcagcagccttaccat 1237  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
Db 1317 RRR 1258  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
QY 1238 ttggaagtgccgaatctccaggggcccagagctgcgcgaggaagaaacccgaag 1297  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
Db 1257 RRR 1198  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
QY 1298 ttccgcgggggagcgggtcgaccccgagcctgcgcgagagagagcccgcggtta 1357  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
Db 1197 RRR 1138  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
QY 1358 acaagcaggcgagcggagcgcctcgcgccacgaggaagagagaggtgcttaag 1417  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
Db 1137 RRR 1078  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
QY 1418 ataggaccagcagagtaaacctcagcccgagcccaagcacag 1462  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...  
Db 1077 RRR 1033  
|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...|||...

RESULT 3  
US-08-930-001-1  
; Sequence 1, Application US/08930001A  
; Patent No. 6281412  
; GENERAL INFORMATION:  
; APPLICANT: MURATA, No. 628141210  
; TITLE OF INVENTION: METHOD FOR PRODUCING OSMOTOLERANT PLANTS  
; FILE REFERENCE: 0230-118P  
; CURRENT APPLICATION NUMBER: US/08/930,001A  
; EARLIER FILING DATE: 1997-09-26  
; EARLIER APPLICATION NUMBER: JAPAN 106819/1995  
; EARLIER FILING DATE: 1995-03-27  
; NUMBER OF SEQ ID NOS: 8  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 2400  
; TYPE: DNA  
; ORGANISM: Arthrobacter globiformis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (361)..(2001)  
; US-08-930-001-1

Query Match 1.4%; Score 47.4; DB 4; Length 2400;  
Best Local Similarity 43.3%; Pred. No. 0.0068;  
Matches 222; Conservative 0; Mismatches 291; Indels 0; Gaps 0;



Db 838877 CCGCGCGCGCGCGCGTTG 838859

## RESULT 6

```

US-09-103-840A-2/c
; Patent No. 6294328
; Application US/09103840A
;
; GENERAL INFORMATION:
; APPLICANT: FLETSCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Owen M.
; APPLICANT: FRASER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007_00
; CURRENT APPLICATION NUMBER: US/09/1
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765

```

```

; AFFILIATION:
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at vari
; OTHER INFORMATION: represent a, t, c
; US-09-103-840A-2

```

Query Match	1.28;	Score 41.2;	DB 4;	Length 4403765;
Best Local Similarity	45.98;	Pred. No. 12;		
Matches 142;	Conservative 0;	Mismatches 168;	Indels 0;	Gaps 0;
QY 2526	cctccggggccagcggcagtcctcatcctgctggccgcagttctcgtgccagatggagaa	2585		
Db 1174714	CGTACCGCGGAGCATCTCACCCAGTGGTTCGCGACATCGCGGTGATCGACATCGACGA	1174655		
QY 2586	gtgggttgaggacatccagatggcattgacctggcggagagagcagcagcccccgcgcc	2645		
Db 1174654	CGSGGTGGCGTTCAACCTCGACACGACGCGTGCAGGAAATTCGGAGCGGGCCGACTA	1174595		
QY 2646	tgaattctcctggccagcagccccctctgacacaaagtccccctgatatgaagccaccgcggctga	2705		
Db 1174594	CCCGGGCCCTCGCGGTGCGAGTCGCTATGTGGTTCGACCGTGCAGGGCATCGCGGCTG	1174535		
QY 2706	ccaggatcagaggatgaacctgacgcctcgcacatcgttgagcagccagcccccga	2765		
Db 1174534	GGATGTGTCACCGGGGACCGATTCGCGCGTGGCCCCACACGGGTGACCATCGACCGGAT	1174475		
QY 2766	ccgcgggaacacaaatggtgcacgtgtgtggcaccgcgaacaccacgcgtctccatggtgga	2825		
Db 1174474	CCTCGGAGCGCGATCACACTCTCTGGGCTACGCGCCGAGACCATCATCGCCGAGAGGG	1174415		
QY 2826	cttcagcatc	2835		
Db 1174414	AGTGACCATC	1174405		

## RESULT

```

US-08-211-682-24
; Sequence 24, Application US/08211682
; Patent No. 5670333
; GENERAL INFORMATION:
;
; APPLICANT:
;
; APPLICANT:
;
; TITLE OF INVENTION:  EXPRESSION OF POLYPEPTIDES IN E. COLI UNDER
;
; TITLE OF INVENTION:  CONTROL OF THE E. COLI MDH-GENE PROMOTER
;
; NUMBER OF SEQUENCES:  25
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE:  Floppy disk
;
; COMPUTER:  IBM PC compatible
;

```

```

? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn.Release #1.0, Version #1.25 (EPO)
? CURRENT APPLICATION DATA:
?   APPLICATION NUMBER: US/08/211.682
?   INFORMATION FOR SEQ ID NO: 24:
?     SEQUENCE CHARACTERISTICS:
?       LENGTH: 1598 base pairs
?       TYPE: nucleic acid
?       STRANDEDNESS: double
?       TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? FEATURE:
?   NAME/KEY: CDS
?   LOCATION: 286..1269
?   FEATURE:
?   NAME/KEY: -10_signal
?   LOCATION: 228..233
?   FEATURE:
?   NAME/KEY: -35_signal
?   LOCATION: 205..210
? US-08-211-682-24
?
```

Query Match	1.28;	Score 40.8;	DB 1;	Length 1598;
Best Local Similarity	45.24;	Pred. No. 0.33;		
Matches 150;	Conservative 0;	Mismatches 182;	Indels 0;	Gaps 0;
Qy 2353	ctctcggggaagggctccagcagcagctgtctctctctgttccaaacagcagctcgtctat	2412		
Db 415	CCCAGGCCATGAAGAGCCCTCGAGGGCGTGTCTATGGAGCTGGAGGACTGCCTTCGCC	474		
Qy 2413	acgagcgggggctgacggctccactcaactcagtttaaaagtcacagggcagctccccctctat	2472		
Db 475	CTGCTTGGGGGCGCTTCGAGGCCACCCAGACCCCAAGGTGGCTTCAAGGACGCCGACTAC	534		
Qy 2473	ggcatcagattgaggagagcgaaagcagtgtygggggtgcceccactgcctgacctccgg	2532		
Db 535	GCCCTCTCTGTGGGGGGCGGCCCGCCGGAAGCGGGGATGGAGCGCGGGAGCTTTGCGAG	594		
Qy 2533	ggccagcgcagtcacatcattggtgcgcaggttctcgtctccgagatgagaagtgggtt	2592		
Db 595	GTAACGCCAAGATCTTCACCGACGAGGGCGGGCCCTCGGCCGAGGTGGCCAAAGAGAC	654		
Qy 2593	gaggacatccagatggccattgacttgcgtggcgagaagagcagcagccccgccttgagttc	2652		
Db 655	GTCAGGTGCTGTTGGTGGGCAACCCCGCCCAACACCCACCGCCCTCATCGCCTACAAGAC	714		
Qy 2653	ctggccagcagccccctgacacaagtcccc	2684		
Db 715	GCCCGCGGCTCAACCCCGGAACTTCACGC	746		

## RESULTS

```

US-08-852-401-1
: Sequence 1, Application US/08852401
: Patent No. 5976836
:
: GENERAL INFORMATION:
:
: APPLICANT: Weber, J. Mark
: APPLICANT: Hessler, Paul E.
: APPLICANT: Larsen, Peter E.
: APPLICANT: Luu, B. Minh
:
: TITLE OF INVENTION: Methods and Compositions for Enhancing
:                               Eyethromycin Production
:
: NUMBER OF SEQUENCES: 5
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Rockey, Milnamow & Katz, Ltd.
: STREET: 2 Prudential Plaza, 180 N. Stetson, Suite
: STREET: 4700
: CITY: Chicago
: STATE: Illinois
: COUNTRY: U.S.A.
:
: ZIP: 60601
:
: COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/852,401  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mueller, Lisa L.  
REGISTRATION NUMBER: 38,978  
REFERENCE/DOCKET NUMBER: FER2159POO30US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-616-5400  
TELEFAX: 312-616-5460  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4776 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-852-401-1

Query Match 1.2%; Score 40.8; DB 2; Length 4776;  
Best Local Similarity 43.9%; Pred. No. 0.56;  
Matches 174; Conservative 0; Mismatches 222; Indels 0; Gaps 0;

QY 2012 tcaagagctcccgccgctggagaaacttctgcagagacttgcagagacttgcagagagtggtgtt 2071  
|||||  
Db 797 TCAACCGCGGGCGGCATCAAGAGGACACCCGCCGCGCTGCGCGCTCAACG 856  
QY 2072 acctaccgttcaaacacttctctctcgtcgcactgcacccggtctcatgcactacaagcagg 2131  
|||||  
Db 857 AGATGCCGACACACCGCGCGGCATAGGTGCGTGCAGCGCACCGGGGTAATCGGCCTGC 916  
QY 2132 tctcgagcggtgtgtcaaacacacaccccgagccagccgacttcaaggactgcagag 2191  
|||||  
Db 917 TGGTCCGCGGAGCTGTCAACCCCGGTGTTCCCGCCCTTCGCGCGAGGCCCTGGAGCGCGCG 976  
QY 2192 ccgctttgcagagatcacggagatggtggcacagctccacggtacagatgatacaagatgg 2251  
|||||  
Db 977 CGTCCGCGGGCGGTACGGCTGCTGCTGTGCAACACCCGCGTGGGATGAGCGAGGAG 1036  
QY 2252 agaattccagaagctgcacgaactcaagaagatttgattgcatgtgacaaattctgtgg 2311  
|||||  
Db 1037 ACTACGTCGGATGCTCATCGCCGCGGCGTGGAGGGCATGGTCTTCGTGTCGCCGAGGA 1096  
QY 2312 ttccgggaaggagttcattcctgtctgtggcagctcgcagactcctcgtgggaagggtctcc 2371  
|||||  
Db 1097 TCGCCCAACACCGAGGCGGAGCGGATCAGCGCGAGCTACTACGAGAAGCTGCTGGCCG 1156  
QY 2372 agcagcgcatgttcttcctgttcaacgacgtctctgc 2407  
|||||  
Db 1157 ACGGCGTGGCATGCTTCTCGTCAACGGCGGCGCGC 1192

RESULT 9  
US-08-949-155-49  
Sequence 49, Application US/08949155  
Patent No. 6271436  
GENERAL INFORMATION:  
APPLICANT: Piedrahita, Jorge A  
APPLICANT: Bazer, Fuller W  
TITLE OF INVENTION: Compositions and Methods for the  
Generation of Transgenic Animal Species  
NUMBER OF SEQUENCES: 51  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: ARNOLD, WHITE AND DURKEE  
STREET: P.O. Box 4433  
CITY: Houston  
STATE: TX

COUNTRY: US  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
FILING DATE: US/08/949,155  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/027,338  
FILING DATE: 11-OCT-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/046,094  
FILING DATE: 09-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Hibler, David W.  
REGISTRATION NUMBER: 41,071  
REFERENCE/DOCKET NUMBER: TAMK:177  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
INFORMATION FOR SEQ ID NO: 49:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4791 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-949-155-49

Query Match 1.2%; Score 40.4; DB 4; Length 4791;  
Best Local Similarity 59.6%; Pred. No. 0.72;  
Matches 68; Conservative 0; Mismatches 46; Indels 0; Gaps 0;  
QY 2569 cggctccagatgagagagtggttgagacatccagatggccattgacctggcgaggag 2628  
|||||  
Db 2866 CGCGCAGGTCTGGAGAGAGAGAGAGCTGGAGCCCTGAGAGGGGAAACGTCCTTGCAGGACG 2925  
QY 2629 agcagacccccccctgaattccttggccagcagcccccttgacaagaagtc 2682  
|||||  
Db 2926 AGTCGAGAGCTTCCCTTAAGCTGTCCTGCCCCCTGCCCCCACCCTCCGACCC 2979

RESULT 10  
US-08-676-967-5  
Sequence 5, Application US/08676967  
Patent No. 5747317  
GENERAL INFORMATION:  
APPLICANT: COLLINS, KATHLEEN  
TITLE OF INVENTION: Human Telomerase  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Science & Technology Law Group  
STREET: 268 Bush Street, Suite 3200  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94104  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/676,967  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard A  
REGISTRATION NUMBER: 36,627





```

QY 2553 cgtggcgcagttctcgttcagatggaagaagtgggttggaggacatccaatgacct 2612
Db 804 GAACGGCGCGTGAAGCGCCCGCCCGCCGCAAGAGCAGCGACCGAGGAGACAG 863
QY 2613 tgacctggcggaagaagcagcagcagcccgccctgagttcctggtccagcagcccccctga 2672
Db 864 CGACCTGGAGGAGAGCGACACATCGACGCGGAGGAGCTGGCCCGACGAGCAGCACCAG 923
QY 2673 caaactgccctgatgaagccaccgcy 2700
Db 924 CACCGAGGAGCAGGAGGACAAAGGCGGTG 951

RESULT 13
US-08-677-734A-8
; Sequence 8, Application US/08677734A
; Patent No. 5871919
; GENERAL INFORMATION:
; APPLICANT: Brant, Steven R.
; APPLICANT: Yun, Chris C.H.
; APPLICANT: Donowitz, Mark
; APPLICANT: Tse, Chung-Ming
; TITLE OF INVENTION: Cloning, Tissue Distribution, and
; TITLE OF INVENTION: Functional Analysis Of The Human Na+/H+ Exchanger Isoform,
; TITLE OF INVENTION: NHE3.
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,734A
; FILING DATE: 10-JUL-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fordis, Jean B.
; REGISTRATION NUMBER: 32,984
; REFERENCE/DOCKET NUMBER: 05387.0043-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2574 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-677-734A-8

Query Match 1.1%; Score 38.6; DB 2; Length 2574;
Best Local Similarity 44.1%; Pred. No. 1.6;
Matches 161; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

QY 2320 agggagttaccgtctgtggcagctcagaaagctctcggggaaggggtccagcagcgc 2379
Db 1762 AGAGAAAATGACGCGTGTGTGCGTGCACATGCAGTCTCTGGACGACGCGGCGGAGC 1821
QY 2380 atgttctctgtttcaacgacgctcgtctatatacagcgcgggggctgacgscctccaat 2439
Db 1822 ATCCGGGACGCGGAGGACATGGTGACGACCACACACGCTACACGATACCTGTACAGCGC 1881
QY 2440 cagtttaagtccacgggcagctccgctctatggcatgacgattgagagagcgaagac 2499

```

Db	3020	1435	3080	1495	3140	1555	3200	1615	3260
Qy	GCCTCTGCCCCCAGGAGCTGATAGAGGAGACATGAAAGGAGTGAAGGCTACCGC	aaactcagcccccgaccccaagcacaggctccctgactgcagctcctcactcttcagag	GAGGAGCTGGAGCGCAGCTGGGCCCTGTACCAGAGAGACGCGAGGCGCGCTGTCCAAG	ctgtctgtgaactcgcaggggggagtgccccctccaaactgacctgtctctcccaacctg	GAGCTCAGCGGCGCAGGCGCGCTGGCGCCGACATGGAGGACGTGCACACCGCTTG	agccccagacaaacagcaggctctcctctgtatcagccgctgtctgaatgcacaggcctgc	GTGCTCTACCGCAGCGAGGTGCACAACTGTGGGCCGAGACCACCGAGGAGCTGCGGAGC	ccccggac	CGCTGCG
Db	3079	1494	3139	1554	3199	1614	3259	1622	3267

RESULT 15  
5352575-6  
Patent No. 5352575  
; APPLICANT: PETROVSKIS, ERIK A.; POST, LEONARD E.; TIMMINS, JAMES G.  
; TITLE OF INVENTION: PSEUDORABIES VIRUS PROTEIN  
; NUMBER OF SEQUENCES: 12  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/513,282  
; FILING DATE: 20-APR-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 100,817  
; FILING DATE: 29-JUN-1987  
; APPLICATION NUMBER: 886,260  
; FILING DATE: 16-JUL-1986  
; APPLICATION NUMBER: 784,787  
; FILING DATE: 04-OCT-1985  
; APPLICATION NUMBER: 801,799  
; FILING DATE: 26-NOV-1985  
; APPLICATION NUMBER: 844,113  
; FILING DATE: 26-MAR-1986  
; SEQ ID NO: 6:  
; LENGTH: 1053  
; 5352575-6

Query Match	1.1%	Score 38.2;	DB 6;	Length 1053;
Best Local Similarity	54.0%;	Pred. NO. 1.3;		
Matches 101;	Conservative 0;	Mismatches 83;	Indels 3;	Gaps 1;
QY 1263	ggccagagctgcgcgagaaagaacgaagtttcgcgcggagacgcgggtgcga	1322		
Db 564	gtgcacgacagctgcgcctgtggaccccgccgagagcgcggtgttcacacgcgcgc	623		
QY 1323	cccgagc---cctgcgcgaggaagccgcgcgggttaacaagcagcgcgagcgcgc	1379		
Db 624	cccgatcgaccagagccgcgcgacacccccgcgcgcgcgcgcgcgcgcgcgc	683		
QY 1380	ctcggggccacgaggaagagagaggtcggttaagatagagaccacgacagataacc	1439		
Db 684	cgagccccgcctccgacgaagagagagacgagaggggggcacgacgcgatgaccc	743		
QY 1440	tcagccc	1446		
Db 744	qatgcc	750		

Search completed: December 6, 2001, 14:32:44  
Job time: 21281 sec

Result No.	Query	Score	Match	Length	DB	ID	Description
c	1	862	25.0	930	10	AL566821	AL566821 AL566821
	2	784.8	22.8	1122	10	BE745268	BE745268 601579755
	3	761.2	22.1	804	11	BG747792	BG747792 602705299
	4	757.2	22.0	865	10	BE260677	BE260677 601154188
	5	755.8	22.0	784	11	BG763918	BG763918 602736888
	6	745	21.6	861	11	BG764061	BG764061 602737071
	7	726.2	21.1	901	11	BG755554	BG755554 602491544
	8	716	20.8	736	11	BF793662	BF793662 602254355
	9	705.4	20.5	898	10	BE745887	BE745887 601573511
	10	704.6	20.5	835	11	BG829192	BG829192 602753244
	11	700.2	20.3	902	11	BE910036	BE910036 601498088
	12	687.6	20.0	721	10	AU132546	AU132546 601432546



```
QY 2499 cgaagtgagggtgccccactgctgacctccgggcccagggcagtcacatcatcgtgc 2558
Db 301 CGAGTGGGGGGTCCCCACATGCTGACCTCCGGGGCCAGCGCATCCATCATCTGCGC 360
QY 2559 cgcagttctcgttcgagatgagagtggttgaggagcatccagatggcattgacct 2618
Db 361 CGCCAGTCTTCGTCGAGATGGAGAGTGGGTGAGGACATCCAGATGGCCATTGACCT 420
QY 2619 ggcggagaagcagcagcccgccctcagttcctggcagcagcccccctgacaacaa 2678
Db 421 GCGGAGAGAGAGCAGCAGCCGCCCTCCTGAGTTCCTGGCCAGCAGCCCTGACAACAA 480
QY 2679 gtccctgatgaagcaccgcggctgaccagagtgagaggtgactgagcgcctcgcg 2738
Db 481 GTCCCTGTATGAAGCACCAGCGCTGACCAGGAGTCAGAGGATGACCTGAGCGCCTCGCG 540
QY 2739 cacatcgttgagcgcagcccgcccgccagcgcgcgcacacaaatggtgcagtgct-ggc 2797
Db 541 CACATCGCTGGAGCGCAGCGCCGCCAGCCGCGGCAACACAATGGTGCACGTGTGCTGGGC 600
QY 2798 accgcaacaccagcgtctccatgg-tggacttcagcatcgagtgaggaatcagttgtct 2856
Db 601 ACGCAACACCAGCGTCTCCATGTTGGACTTCAGCATCGCAGTCGAGAAATCAGTTGTCT 660
QY 2857 ggaacctcgtgaggaatcctcaaaacgaaacgggtggcagaagctgt-gggtggtgtt 2915
Db 661 GGAACCTGCTGAGGAATTCAAAAACAGCAACGGGTGGCAGAAAGCTGTGGGTGGTGT 720
QY 2916 ccaaaactcctcgtctcttctacaaatcacacaggaacatcctccctgtcagcct 2975
Db 721 CACAAACTTCTGCCTGCTTCTTACAAATCACACACAGGAC-ATCATCCCTCTGCCAG- 776
QY 2976 gcctcgtcgtcgtactcgtcaccatccctcgtgctcgagaaacatccagaagacta 3035
Db 777 --CTGCTCTGGTAAATCGTACATACCTCTGAGTCGGAGAACATCCAGAAAGACTA 834
QY 3036 cgt-gtcaagtcgactcctcaagtcacagtcactactactactcagggcggaagcagta 3094
Db 835 GGTGGTTCAGGCGGACTTCTATGCTCTCAGGTCTACTTACTTACAGG-GGAAAGCGAGACC 893
QY 3095 cgttcgaaggtgagtggaagtgatccgcagtcgacccagctct 3138
Db 894 CGTCCGACAGCGGTGGGAATAGTACCCGCGAGGCAACAGCGCT 937

RESULT 3
BG747792
LOCUS BG747792
DEFINITION BG747792 804 bp mRNA EST 15-MAY-2001
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
mRNA sequence.
ACCESSION BG747792
VERSION BG747792.1 GI:14058445
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 804)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaaps-remail.nih.gov
Tissue procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LCM1675 row: j column: 24
High quality sequence stop: 796.
```

```
FEATURES
Source
Location/Qualifiers
1..804
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4841975"
/tissue_lib="NIH_MGC_43"
/tissue_type="normal pigmented retinal epithelium"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library. |"
BASE COUNT 225 a 191 c 209 g 179 t
ORIGIN
Query Match 22.1%; Score 761.2; DB 11; Length 804;
Best Local Similarity 99.1%; Pred. No. 5.3e-181;
Matches 797; Conservative 0; Mismatches 3; Indels 4; Gaps 3;
QY 487 ggcaggttgacgtgtaatagcaccagcagctctctgtattcacacattgtgcaatct 546
Db 2 GGCAGGTTGACGTGTAATGACACGACGCGACTCTCTGATTTCACACATTGTGCAATCT 61
QY 547 ggaattgggatttgatgaagccttggacagagacacttagcaaaaataatacata 606
Db 62 GAGATTGGGATTTTGTATGAAGCCTTGGACAGAGACACTTAGCAAAAAATAATACATA 121
QY 607 cctcagcaagacgcactagaggacaaaatcgtggaatttcaccataaccacattggacaa 666
Db 122 CCTCAGCAAGACGCACCTAGAGGACAAAATCGTGAATTTCCACCATTAACCATTTGGACAA 181
QY 667 acaccgcgaatcagatttccagctcctagagattgcccgtcggctagagatgtagga 726
Db 182 ACACCAAGCAGAAATCAGATTTCAGCTCTTAGAGATTGCCGTCGGCTAGAGATGTATGA 241
QY 727 atccggttcaccccgccaggaagcaggaagcagcaagaatcaatctggcgttgcaac 786
Db 242 ATCCGTTGACCCCGCCAGGACAGGGAAGCGACGAGATCAATCTGGCCGTGGCAAC 301
QY 787 acgggaattctagtttccaggtttcactaaagatcaatgccttcaactgggccaaggtg 846
Db 302 ACGGAATTCCTAGTGTTCAGGTTTCACTAAGATCAATGCTTCACTGGGCAAGGTG 361
QY 847 cggaaactgagcttcaagagaagcgccttctcacaagctccggccagatgcaaatagt 906
Db 362 CGGAAGCTGAGCTTCAAGAGGAAGCGCTTTCTCATCAAGCTCCGGCCAGATGCCAATAGT 421
QY 907 gcgtaccagatacccttgaattcctgtagtgccagtcgggatttctgcaagtcctctg 966
Db 422 GCGTACAGGATACCTTGGAAATTCCTGATGCGCAGTCGGGATTTCTGCAAGTCTCTTCTG 481
QY 967 aaaaactgtgtgaacatcatgccttctttagacttttgaagagcccaaccacaaagccc 1026
Db 482 AAAATCTGTGTGAACATCATGCTTCTTTAGACTTTTGAAGAGCCCAACCAAGGCC 541
QY 1027 aagccgcctctttagccgggggtcatcatcttcggttcagttcggactcagaagcag 1086
Db 542 AAGCCCGTCTCTTTAGCCGGGGTTCATCAATTTTCGTTTCAGTTCGGAGTCAGAACAG 601
QY 1087 gtctcgactatgtt--aagaagaggagacataaagaagtgagtttgaagaagacaca 1144
Db 602 GTTCTCGACTATGTTACACGAAGGAGGACATGAAGAAGTTCAGTTTGAAGAAGCACA 661
QY 1145 gcaagattcattctccggagccttgccttcacagctacagaaactgaatt-cggaagtg 1203
Db 662 G-AAGATTTCATTCTATCCGGGCGCTTGTCTTACACGCTACAGAACTGAATTCGGAAGTG 720
QY 1204 ctggagcagtcctcagcagagaccacgccttacaatttggagaaggtgccgaatctccaggg 1263
```

```
|||||
Db 721 CTGAGCAGTCTCAGCAGACACAGCTTACATTTGAGAAGTGCCGAATCTCCAGG 780
|||||
QY 1364 ggcagagctgccggcgaggaag 1287
|||||
Db 781 GCACAGAGTCCCGCGGAGGAAG 804
|||||

RESULT 4
LOCUS BE260677 865 bp mRNA EST 26-OCT-2000
DEFINITION 601154188F1 NIH_MGC_19 Homo sapiens cDNA clone IMAGE:3510381 5',
mRNA sequence.
ACCESSION BE260677
VERSION BE260677.1 GI:9132166
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 865)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM194 row: o column: 22
High quality sequence stop: 730.
FEATURES
source
1..865
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3510381"
/clone_lib="NIH_MGC_19"
/tissue_type="neuroblastoma"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: brain; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGACGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."
BASE COUNT 207 a 236 c 255 g 167 t
ORIGIN

Query Match 22.0%; Score 757.2; DB 10; Length 865;
Best Local Similarity 97.3%; Pred. No. 5.5e-180;
Matches 823; Conservative 0; Mismatches 18; Indels 5; Gaps 5;

QY 719 tgtatgaattccggtgtgaccggccgaagacagaggaaggaagacagatcaattctggccg 778
|||||
Db 1 TGTATGAATCCGGTTGACCCGCGCAAGACAGGACAGGACGAAGATCAATCTGGCCG 60
|||||
QY 779 ttgccacacgggaattcaggtttcaggtttcactaagatcaatgccttcaactggg 838
|||||
Db 61 TTGCCACACGGGAATCTAGTGTTCAGGGTTTCACTAGATCAATGCCCTTCAACTGGG 120
|||||
QY 839 ccaaggtgcgaagctgagcttaagaggaagcgctttctcatcaagctccggccagatg 898
|||||
Db 121 CCAAGGTGCGGAAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGCTCCGCCAGATG 180
|||||
QY 899 ccaatagtcgtaccaggatacattggaattctgtgatggtcgagtcgggattctgcaagt 958
|||||
Db 181 CCAATAGTCGTACCAAGGATACCTTGGAAATTCCTGATGGCCAGTCGGGATTTCTGCAAGT 240
|||||
```

Qy	1890	tgcccaaatcagagattaccacaagaatcggcgatgtcatgc-tgaagaacatttcaggcca	1948
Db	722	TGCCAAATCAGAGATTACCAAGAATCGCGGATGTGTCATGCTTTGACGAACATTACGCCCA	781
Qy	1949	tga	1951
Db	782	TGA	784
RESULT	6		
LOCUS	BG764061	861 bp mRNA	EST 15-MAY-2001
DEFINITION	60237071F1 NIH_MGC_49 Homo sapiens cDNA clone IMAGE:4862417 5',		
ACCESSION	BG764061	mRNA sequence.	
VERSION	BG764061.1	GI:14074714	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 861)		
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/		
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)		
JOURNAL	Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgabbs-femail.nih.gov Tissue procurement: ATCC/DCTD/DTP cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCMI722 row: n column: 18 High quality sequence stop: 817. Location/Qualifiers 1. .861 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:4862417" /clone_lib="NIH_MGC_49" /tissue_type="melanotic melanoma, high MDR (cell line)" /lab_host="DH10B (phage-resistant)" /notes="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACAG(G). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies).. Note: this is a NIH_MGC Library. I"		
BASE COUNT	198 a	245 c	270 g 147 t 1 others
ORIGIN			
Query Match	21.6%;	Score 745;	DB 11; Length 861;
Best Local Similarity	96.9%;	Pred. No. 6.6e-177;	
Matches	791;	Conservative	0; Mismatches 21; Indels 4; Gaps 3;
Qy	1910	aaagaatcggcgatgtcatgtcgtgaagaacattcagggcattcaggacacctggcgctcacc	1969
Db	2	AAAGAATCGGGATGTCTGCTGAAGAACAATTCAGGCGCATGAAGCACCTGGCGGCTCACC	61
Qy	1970	tgtggaagcacagcgaggccttgaggcccttgagagatggaatcgaagatcccgccggcg	2029
Db	62	TGTGGAAGCACACGAGGAGCCCTTGAGGCCCTTGAGGAATGGAATCAAGAGCTCCCGCGGC	121
Qy	2030	tggaagaacttctcagagactttgagctgcagaagtggtgtacctaccgctcaacacct	2089
Db	122	TGGAGAACTTCTGCAGAGACTTTGAGCTGCAGAAGGTGTTTACCTACCGCTCAACACCT	181





Db	722	GACAATCTGTGGTCCGGGAACGGAGTTCCATCCGGTCTGGGCAGCTTCAAGCAAGC	781
Qy	2352	-gctctcgggaaagggtctccagcagcgcatgttcttctctgttcaacgagctct-gcta	2409
Db	782	TCCTCGGGGAAGGACTCCACGACGCGCATGCTTCCCTGTGTCACGAAGCTCTGGCTT	841
Qy	2410	tacacgacggggggctgacggcctccaatacagtttaaaagtcacggcgagctcc	2464
Db	842	TACACGACCGGGGCTGACAGCGCTCCAATCCAGCTTAAAGCCAAGGCGAGTCCC	896
RESULT 8			
LOCUS	BF793662	736 bp	EST 12-JAN-2001
DEFINITION	602254351F1 NIH_MGC_84 Homo sapiens cDNA clone IMAGE:4346509 5', mRNA sequence.		
ACCESSION	BF793662		
VERSION	BF793662.1	GI:12098716	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 736)		
TITLE	NIH-MGC http://mgc.nci.nih.gov/.		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
COMMENT	Contact: Robert Strausberg, Ph.D. Email: cgapbs-r@mail.nih.gov Tissue procurement: ATCC cDNA Library Preparation: Life Technologies, Inc. DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLAM9968 row: f column: 14 High quality sequence stop: 710.		
FEATURES	Location/Qualifiers		
source	1..736		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:4346509"		
	/clone_lib="NIH_MGC_84"		
	/tissue_type="adrenal cortex carcinoma, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: adrenal gland; Vector: pCMV-SPORT6; Site:1-		
	Noti; Site:2: SalI; Cloned unidirectionally; oligo-dr		
	primed. Average insert size 1.229 kb. Library enriched for		
	full-length clones and constructed by Life Technologies.		
	Note: this is a NIH_MGC Library."		
BASE COUNT	167 a	228 c	209 g 132 t
ORIGIN			
Query Match	20.8%; Score 716; DB 11; Length 736;		
Best Local Similarity	99.2%; Pred. No. 1.3e-169;		
Matches	730; Conservative	0; Mismatches	5; Indels 1; Gaps 1;
Qy	2397	cgagctcctgctatcacgacgacgggggtgacggcctccaatcagtttaagtcacgg	2456
Db	1	CGAGCTCTGCTATACAGACGCGGGGCTGACGGCTCCAATCAGTTAAAGTCCACGG	60
Qy	2457	gcagctccgctctatggcatgacgattgaggagagcgaagacgagtgagggggtgcccca	2516
Db	61	GCAGCTCCGCTCTATGGCATGACGATGTAGGAGAGCGAAGACAGAGTGGGGGTGCCCA	120
Qy	2517	ctgcctgacctccggggccagcgagtcacatcgtgcccacgagttctcggtccga	2576
Db	121	CTGCTGACCCCTCCGGGGCCAGCGGCGAGTCCATCATCTGTGGCCCGCAGTTCTCTGGTCCGA	180
Qy	2577	gatggagaagtggattgagacatccagatggccttgacctgcccgggagaagacgacg	2636



```
Db 301 CGCCAGGAAGAACCGAAGGTTTCGCCGGGGAGCGGGTCCACCCAGCCGCTCGG 360
Qy 1336 ccgaggagaagcccgcggggtaacaagcagcgagcgagccgctcgccgccccacgag 1395
Db 361 CCGAGGAGAAGCCCGCGGGTAACAAGCAGCGGAGCGGCGCTCGCGCCACCGAG 420
Qy 1396 gaagaggaggagtgctt-aagtagtagaccagcagagtagtaaacctcagccccagacc 1454
Db 421 GAAGAGGAGGAGTGGTTCAAGGATAGACCCAGACAGAGTAACCTCAGCCCCCGACCC 480
Qy 1455 aagcacaggtccctgactggcagctccacgttcccgagctgtctgtaactcgagg 1514
Db 481 AAGCACAGCTCCCTGACTGGCAGTCCCTACCTTTCCGAGCTGTGTGAACCTCGCAGG 540
Qy 1515 gggagtgccccctgcacacgtgacgttctctcccaacctgagccccgacacacagcagc 1574
Db 541 GGGAGTGGCCCTGCGCAACGTGACCTTGTCTCCAGACTGAGCCCCGAAACCAAGCAGCC 600
Qy 1575 ctctcccttgatcagccgctgc-tgaatgaccagcctgcccccgagcagcagatgagg 1633
Db 601 CTCCTCCCTTGATCAGCCCGCTGCTTGAATGACCAAGGCTGCCCCGCGACGACGATGAGG 660
Qy 1634 atgagggcggaggaagagattcccaactgataaagcgtacttca-tagctaaagaaagt 1692
Db 661 ATGAGGCGCGGAGGAGAGATTCCTCCACTGATAAAGCGTACTTCAATTAGCTAAGGAATG 720
Qy 1693 tctaccacgagcgaac---atatggaagatctcgaag-ttatcacttctgtgttca 1748
Db 721 TCTACCACCGAGCAACCATATTCTGAAGGATCTCGAAGTTTATCACTTCGCGGTTTACG 780
Qy 1749 gagcacagtgaagaaaggacgcatccaccccggaagcac---tgaaaagtctcata 1800
Db 781 AAGCACAGTGAAGCAAGAGGAGCGCATGCCGGAAGACACTTGAAGAAGTCTCAT 835

RESULT 11
BE910036 601498085F1 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:3900021 5',
LOCUS mRNA sequence. EST 20-OCT-2000
DEFINITION BE910036
ACCESSION BE910036
VERSION BE910036.1 GI:10406227
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 902)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: csaabs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LHAM9698 row: j column: 22
High quality sequence stop: 750.
Location/Qualifiers
1. 902
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3900021"
/clone_lib="NIH_MGC_70"
/tissue_type="epitheloid carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
```

```
BASE COUNT 231 a 248 c 272 g 151 t
ORIGIN
Query Match 20.3%; Score 700.2; DB 11; Length 902;
Best Local Similarity 95.8%; Pred. No. 1.3e-165;
Matches 752; Conservative 0; Mismatches 15; Gaps 3;
Qy 1191 gaattcgaagtgtctggagcagctctcagcagagcaccagccttaccatttggagaaggtagc 1250
Db 1 GAATTCGGAAGTGTCTGGAGCAGTCTCAGCAGAGCACCAGCCTTACATTTGGAGAAGGTGC 60
Qy 1251 cgaatctccagggggcccagagctgcggcgaggaaagaaacgaagggtttccgcgcgggga 1310
Db 61 CGAATCTCCA-GGGGCCAGAGCTGCGCGAGGAAGAAACCGAAGGTTCGCGCGGGGA 119
Qy 1311 gccggggtcgcacccagcagcctcgccgagagagagcccggttaacaacagcagcgga 1370
Db 120 GCCGGGGTGGCACCCGAGCCCTGCGCGGAGGAGAGCCCGCGGGTAAACAAGCAGCGCGA 179
Qy 1371 cggagccgcctcggcgccacgagagagagagagaggtcggttaaggatagagaccagca 1430
Db 180 CGGAGCCGCTCGGCGCCACGAGGAGAGGAGGAGGTCTGTTAAGGATAGGACCCAGCA 239
Qy 1431 gagtaaacctcagcccccgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1490
Db 240 GAGTAAACCTCAGCCCGCGAGCAAGCAAGCAGCTCCCTGACTGGCAGTCTCTCACTTTC 299
Qy 1491 cgagctgtctgtaactcgagggggagtgccctgcgaagcagcagcagcagcagcagcagc 1550
Db 300 CGAGTGTCTGTGAACCTCGAGGGGGAGTGGCCCTGCAACGTCACCTTCTCTCCCAA 359
Qy 1551 cctgagccccacacacagcagcagcctctcccttgatcagccgctgctgaatgacagcagc 1610
Db 360 CTTGAGCCCGGACCAACAGCAGGCTCTCCCTTGATCAGCCCGCTGCTGAATGACCAGC 419
Qy 1611 ctgccccggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1670
Db 420 CTGCCCCCGGAGGACGATGAGGATGAGGGCGGAGGAAGAGATTCCCAACTGATAAAGC 479
Qy 1671 gtacttcagtaaggagagtgctaccacccagcagcagcagcagcagcagcagcagcagc 1730
Db 480 GTACTTCATAGCTAAGGAAGTGTCTTACCACCGAGCGACATATCTGAAGGATCTCGAAGT 539
Qy 1731 tatcacttcgtgttccagagcacagtgagcaaaagagcagcagcagcagcagcagcagcagc 1789
Db 540 TATCACTTCGTGGTTCAGAGCAGCAGTGAAGCAAGAGAGAGCGCCATGCCGGAAGCAGTGA 599
Qy 1790 aaa---gtctcatattcccgaattttgaaccttttcacaaatttcatactaatatttctca 1846
Db 600 AAAAGTCTCATATTCCCGAATTTTGAACCTTTGCACAAATTTTCATACTAATTTCTCTCA 659
Qy 1847 aggaattgagcaacgacttgcctgtgtgggaaggcgctcaaatgcccaaatcagagatt 1906
Db 660 AGGAATTTGAGCAACGACTTGCCTGTGGGAAAGCCGCTCAATGCCAAATCCCAATCAGAGATT 719
Qy 1907 accaagaatcggcagctatcgtcgtgaagacattcagggcagcagcagcagcagcagcagc 1966
Db 720 ACCAAGAATCGCGATGTCATGCTGAAGAACCATTTACGCCATAAAGCACCCGCGGGGTCC 779
Qy 1967 acctg 1971
Db 780 CTGTG 784

RESULT 12
AUI32546
LOCUS AUI32546 NT2RP4 Homo sapiens cDNA clone NT2RP4000050 5', mRNA
DEFINITION AUI32546
ACCESSION AUI32546
VERSION AUI32546.1 GI:10992900
```

KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 721)  
Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y., Saito,K., Yamamoto,J.,  
Nishikawa,T., Nakamura,Y., Nagai,T., Sugano,S., Masuho,Y. and  
Isogai,T.  
TITLE HRI human cDNA project (Ota,T., Sugiyama,T., Ishii,S., Suzuki,Y.,  
Saito,K., Yamamoto,J., Nishikawa,T., Nakamura,Y., Nagai,T., Sugano  
,S., Masuho,Y., Isogai,T.)  
JOURNAL Unpublished (2000)  
COMMENT Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3951  
Fax: 81-438-52-3952  
Email: genomics@hri.co.jp  
HRI human cDNA project: 5'- & 3'-end one pass sequencing: Helix  
Research Institute; cDNA library construction: Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.  
FEATURES  
Source Location/Qualifiers  
1..721  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="NT2RP400050"  
/cell\_lib="NT2RP4"  
/cell\_type="teratocarcinoma"  
/note="Vector: pME18SPL3; mRNA from NT2 neuronal precursor  
cells after 2-weeks retinoic acid (RA) induction"  
BASE COUNT 183 a 180 c 214 g 141 t 3 others  
ORIGIN  
Query Match 20.0%; Score 687.6; DB 10; Length 721;  
Best Local Similarity 98.8%; Pred. No. 1.9e-162;  
Matches 712; Conservative 0; Mismatches 7; Indels 2; Gaps 2;  
QY 709 cggctagagatgctggaatccggttgccaccggcccaagagcaggaagcagaagtc 768  
DB 1 CGGCTAGAGATGATGGAATCCGTTGTCACCCCGCCCAAGCAGAGGAAGCAGGAATC 60  
QY 769 aatctggccgttgcccaacacgggaattctagtgtttcagggtttcactaatgacc 828  
DB 61 AATCTGGCCGTTGCCAACACGGGAATCTAGTGTTCAGGGTTTCACATGAATCAATGCC 120  
QY 829 ttcaactggcccaagtgccggaagctgagcttcaagaggaagcgctttctcatcaagctc 888  
DB 121 TTCAACTGGGCCAAGCTGCGGAAGCTGAGCTTCAAGAGGAAGCGCTTCTCATCAAGCTC 180  
QY 889 cggccagatgccaatagtcgtaccaggtatcatttggaattcctgattgcccagtcggat 948  
DB 181 CGGCCAGATGCCAATAGTCGCTACCAAGATACCTTGGAAATCTGTATGTCGACGCGGAT 240  
QY 949 ttctgaagtcctctggaataatctgttgaaatcatgctctctttgactttttgaa 1008  
DB 241 TTTCTGAAGTCCTCTGGAATAATCTGTGTGAACATCATGCCCTCTTTAGACATTTTGA 300  
QY 1009 gaccccaacccaaagcccaagcccgctcttttagccgggggtcatcttcggttcagt 1068  
DB 301 GAGCCCAACCAAGCCCAAGCCCGCTCTTTAGCCGGGGTTCATCATTCGGTTCAGT 360  
QY 1069 ggtcggactcagaagcaggttcttcgactatgtttaaagaaggagacataagaaggtgcag 1128  
DB 361 GGTCCGACTCAGAAGCAGGTTCTCGACTATGTTAAAGAAGGAGGACATAGAAGGTGCAG 420  
QY 1129 ttgaaaggaagcagaagatttcttccgagccttcttcacagcctacagaa 1188  
DB 421 TTTGAAAGGAAGCACAGCAAGATTCATCTATCCGGAGCGCTTGCCTTCACAGCGCTACGGAA 480

QY 1189 ctgaattcgaagtgctggagcagctctcagcagcagcaccagccttacatttgagaaggt 1248  
DB 481 CTGAATTCGAAGTGTGGAGCAGTCTCAGCAGCAGCACCAGCCTTACATTTGGAGAAGGT 540  
QY 1249 gccgaatctccggggggccagagctccggcgaggagaaagaaagcgaaggttttcgcgcggg 1308  
DB 541 GCGGATCTCCAGGGGGCCAGAGCTCCGCGGAGGAAGAACCGAAGGTTTTCGCGCGG 600  
QY 1309 gaa-ccgaggttcgacccagccctgcgcgaggaagaagcccgccggttaacaagcagcc 1367  
DB 601 GAGCCCGGGGTGCGACCCAGCCCTCGCCGAGGAGAACCCCGCGGTAAACAGCAGGC 660  
QY 1368 ggacggagcgcctcgcccgccacgaggaagaggaggaggtcg-taaaggatagacccc 1426  
DB 661 GGACGGAGCCGCTNGCGCCACGAGGAGGAAGANGANGTCTGTTAAGGATAGGACCC 720  
QY 1427 a 1427  
DB 721 A 721

RESULT 13  
BG323704  
LOCUS 602421995F1 NIH\_MGC\_14 Homo sapiens cDNA clone IMAGE:4560191 5',  
DEFINITION mRNA sequence.  
ACCESSION BG323704  
VERSION BG323704.1 GI:13130141  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 942)  
AUTHORS NIH-MGC http://mgc.mci.nih.gov/  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: DCTD/DTP  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLCM1268 row: m column: 24  
High quality sequence stop: 760.

FEATURES  
Source Location/Qualifiers  
1..942  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:4560191"  
/cell\_lib="NIH\_MGC\_14"  
/tissue\_type="renal cell adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: kidney; Vector: pOTB7; Site:1: XhoI; Site\_2:  
EcoRI; cDNA made by oligo-dT priming. Directionally  
cloned into EcoRI/XhoI sites using the following 5'  
adaptor: GGCACGAG(G). Size-selected >500bp for average  
insert size 1.8kb. Library constructed by Ling Hong in  
the laboratory of Gerald M. Rubin (University of  
California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."  
BASE COUNT 237 a 264 c 286 g 155 t  
ORIGIN

Query Match 19.7%; Score 679.6; DB 11; Length 942;  
Best Local Similarity 96.8%; Pred. No. 2.1e-160;  
Matches 736; Conservative 0; Mismatches 19; Indels 5; Gaps 4;

D <sub>b</sub>	538	GCCCAAGCCCCCTCCTTTTAGCCGGGGGTATCATTTTCGGTTCAGTGCTCGACTCAGAA	597
Q <sub>y</sub>	1083	qcaggtttctgcactatggtt-aaagaagaggagacataagaagdggtgcagtttgaaaagggaagc	1141



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 6, 2001, 08:51:43 ; Search time 72.4 Seconds  
(without alignments)  
93.103 Million cell updates/sec

Title: US-09-555-342A-2\_COPY\_764\_854

Perfect score: 468

Sequence: 1 GSLSKLGGKLGQRMFFLFN.....IVAASSRSEKWKVEDIQMA 91

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*

1: /SID88/gcgdata/geneseq/geneseq/AA1980.DAT:\*  
2: /SID88/gcgdata/geneseq/geneseq/AA1981.DAT:\*  
3: /SID88/gcgdata/geneseq/geneseq/AA1982.DAT:\*  
4: /SID88/gcgdata/geneseq/geneseq/AA1983.DAT:\*  
5: /SID88/gcgdata/geneseq/geneseq/AA1984.DAT:\*  
6: /SID88/gcgdata/geneseq/geneseq/AA1985.DAT:\*  
7: /SID88/gcgdata/geneseq/geneseq/AA1986.DAT:\*  
8: /SID88/gcgdata/geneseq/geneseq/AA1987.DAT:\*  
9: /SID88/gcgdata/geneseq/geneseq/AA1988.DAT:\*  
10: /SID88/gcgdata/geneseq/geneseq/AA1989.DAT:\*  
11: /SID88/gcgdata/geneseq/geneseq/AA1990.DAT:\*  
12: /SID88/gcgdata/geneseq/geneseq/AA1991.DAT:\*  
13: /SID88/gcgdata/geneseq/geneseq/AA1992.DAT:\*  
14: /SID88/gcgdata/geneseq/geneseq/AA1993.DAT:\*  
15: /SID88/gcgdata/geneseq/geneseq/AA1994.DAT:\*  
16: /SID88/gcgdata/geneseq/geneseq/AA1995.DAT:\*  
17: /SID88/gcgdata/geneseq/geneseq/AA1996.DAT:\*  
18: /SID88/gcgdata/geneseq/geneseq/AA1997.DAT:\*  
19: /SID88/gcgdata/geneseq/geneseq/AA1998.DAT:\*  
20: /SID88/gcgdata/geneseq/geneseq/AA1999.DAT:\*  
21: /SID88/gcgdata/geneseq/geneseq/AA2000.DAT:\*  
22: /SID88/gcgdata/geneseq/geneseq/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	468	100.0	395	21	Human pancreatic c
2	468	100.0	1045	20	Human chondrocyte-
3	468	100.0	1045	21	Human cytoskeleton
4	152.5	32.6	432	22	Human protein sequ
5	122	26.1	766	21	Rat actin-binding
6	118.5	25.3	183	22	Human polypeptide
7	118.5	25.3	383	22	Human polypeptide
8	118.5	25.3	404	22	Human polypeptide
9	112.5	24.0	279	22	Human protein sequ
10	111.5	23.8	249	22	Human polypeptide
11	111.5	23.8	249	22	Human protein sequ

12	111.5	23.8	249	22	AAB64412	Amino acid sequenc
13	85.5	18.3	762	22	AAB94398	Human protein sequ
14	84.5	18.1	394	18	AAW26726	Human integrin reg
15	84.5	18.1	394	20	AAW26928	Human integrin reg
16	84.5	18.1	394	22	AAE00737	Human integrin reg
17	81.5	17.4	201	22	AAB93655	Human protein sequ
18	81	17.3	245	22	AAM39355	Human polypeptide
19	81	17.3	246	22	AAW41141	Human polypeptide
20	81	17.3	301	22	AAB94156	Human protein sequ
21	81	17.3	408	22	AAB95198	Human protein sequ
22	80.5	17.2	505	19	AAW61027	Murine guanine nuc
23	80.5	17.2	535	19	AAW61028	Murine guanine nuc
24	80.5	17.2	554	19	AAW61026	Murine guanine nuc
25	80.5	17.2	814	14	AAW43578	Peptide which modu
26	79.5	17.0	710	20	AAW80995	Amino acid sequenc
27	79.5	17.0	1244	21	AAW68825	Human colon carcin
28	77	16.5	619	22	AAB97025	Human integrin reg
29	75.5	16.1	399	18	AAW33417	Human integrin reg
30	75.5	16.1	399	20	AAW26927	Human integrin reg
31	75.5	16.1	399	22	AAE00736	Human integrin reg
32	75.5	16.1	399	22	AAE00739	Human integrin reg
33	75	16.0	417	22	AAW73491	Human cytohesin-2
34	74.5	15.9	399	22	AAE00740	Human integrin reg
35	72.5	15.5	378	22	AAW42353	Human polypeptide
36	71	15.2	129	19	AAW54285	Human modified cyt
37	71	15.2	263	18	AAW18783	Cytohesin 2 encode
38	71	15.2	398	18	AAW18782	Cytohesin 1. Homo
39	71	15.2	398	22	AAE00738	Human B2-1 protein
40	70.5	15.1	399	19	AAW57707	Mouse GRP protein
41	69.5	14.9	537	18	AAW37045	Murine alpha-(2) s
42	68	14.5	1227	22	AAW65637	Novel protein kina
43	68	14.5	1289	20	AAW27163	Peptide seq ID No:
44	68	14.5	1289	21	AAW56781	Human Trad protein
45	67.5	14.4	487	22	AAU00628	Novel human protei

#### ALIGNMENTS

RESULT 1

AAB54227

ID AAB54227 standard; Protein; 395 AA.

XX AC AAB54227;

DT 09-MAR-2001 (first entry)

XX DE Human pancreatic cancer antigen protein sequence SEQ ID NO:679.

XX KW Human; pancreas; pancreatic cancer; pancreatic cancer antigen;

XX KW detection; diagnosis; identification; cytosolic; neuroprotective;

XX KW nontropic; immunomodulatory; relaxant; contraceptive; gynaecological;

XX KW antinflammatory; cardiant; gene therapy; chromosome mapping;

XX KW linkage analysis; tissue identification; tissue typing; forensic;

XX KW neural; immune system; muscular; reproductive; gastrointestinal;

XX KW pulmonary; cardiovascular; renal; proliferative.

XX OS Homo sapiens.

XX PN WO200055320-A1.

XX PD 21-SEP-2000.

XX PF 08-MAR-2000; 2000WO-US05989.

XX PR 12-MAR-1999; 99US-0124270.

XX PA (HUMA-) HUMAN GENOME SCI INC.

XX PI Rosen CA, Ruben SM;

XX DR WPI: 2000-579444/54.

DR N-PSDB; AAC98992.



XX New nucleic acid that is a pancreatic cancer antigen for preventing,  
 PT treating, or ameliorating a medical condition, particular pancreatic  
 PT cancer, or for use in assays for diagnosing a pathological condition -  
 XX  
 PS Claim 11; Page 1115-1116; 1379pp; English.  
 CC  
 CC AAC98773 to AAC99231 encode the human pancreatic cancer associated  
 CC proteins, called pancreatic cancer antigens, given in AAB54008 to  
 CC AAB54466. The human pancreatic cancer antigens have cytostatic,  
 CC neuroprotective, neurotropic, immunomodulatory, relaxant, contraceptive,  
 CC gynaecological, cardiant and antiinflammatory activities, and can be used  
 CC in gene therapy. The polynucleotide and proteins can be used for  
 CC preventing, treating, or ameliorating a medical condition or in assays  
 CC for diagnosing a pathological condition or a susceptibility to one in a  
 CC subject. Binding partners to the proteins and the activity of the  
 CC proteins can be identified. The pancreatic cancer antigens can be used to  
 CC detect, treat or prevent pancreatic disorders, especially cancer.  
 CC Agonists and antagonists to the antigens can be screened for. The  
 CC pancreatic cancer antigen polynucleotides can be used to design nucleic  
 CC acid hybridisation probes that can be used in chromosome mapping, linkage  
 CC analysis, tissue identification and/or typing and a variety of forensic  
 CC and diagnostic methods. The proteins can be used to generate antibodies  
 CC which are used to purify, detect and target the polypeptides, including  
 CC both in vivo and in vitro diagnostic and therapeutic methods. The  
 CC proteins can be used to treat or prevent neural, immune system, muscular,  
 CC reproductive, gastrointestinal, pulmonary, cardiovascular, renal or  
 CC proliferative disorders. AAC99232 to AAC99240 and AAB54467 represent  
 CC sequences used in the exemplification of the present invention.  
 XX  
 SQ Sequence 395 AA;

Query Match 100.0%; Score 468; DB 21; Length 395;  
 Best Local Similarity 100.0%; Pred. No. 8.5e-56;  
 Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSLSKLSGKGLQRMFFLNDVLLYTSRGLTASNQPKVHGQLPLYGMTIESEDEWGVPH 60  
 Db 220 gsisklskgqlqrmfflndvlltsrgltasnkqkvhgqlplygmtieesedewgvph 279  
 QY 61 CLTLRGQRQSIIVAAASRSRSEMEKWWEDIQMA 91  
 Db 280 cltlrgqrqsiivaaasrsrsemekwwediqma 310

RESULT 2  
 AAY07482  
 ID AAY07482 standard; Protein; 1045 AA.  
 XX  
 AC AAY07482;  
 XX  
 DT 17-AUG-1999 (first entry)  
 XX  
 DE Human chondrocyte-derived protein CDEP.  
 XX  
 KW Differentiation; human; foetal; chondrocyte; ezrin-like domain; cancer;  
 KW Dbl1 homology domain; pleckstrin homology domain; rheumatoid arthritis;  
 KW drug.  
 OS Homo sapiens.  
 XX  
 PN W09928458-A1.  
 XX  
 XX 10-JUN-1999.  
 PD  
 XX 27-NOV-1998; 98WO-JP05348.  
 PF  
 XX 27-NOV-1997; 97JP-0342060.  
 PR  
 XX (CHUS ) CHUGAI SEIYAKU KK.  
 PA  
 XX Kato Y, Kawamoto T, Koyano Y;  
 PI

XX WPI: 1999-371117/31.  
 DR N-PSDB; AAX79183.  
 XX  
 PT Protein CDEP expressed in differentiated chondrocytes, and gene  
 PT encoding it  
 XX  
 PS Claim 2; Fig 1; 59pp; Japanese.  
 CC  
 CC This sequence represents a protein (CDEP) expressed in differentiated  
 CC human foetal chondrocytes, which contains an ezrin-like domain, a Dbl  
 CC homology (PH) domain and a pleckstrin homology (PH) domain. The encoding  
 CC nucleic acid or protein can be used in the investigation and treatment of  
 CC cancers and arthritic diseases (including chronic rheumatoid arthritis),  
 CC or for screening of candidate anticancer drugs.  
 XX  
 SQ Sequence 1045 AA;

Query Match 100.0%; Score 468; DB 20; Length 1045;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-55;  
 Matches 91; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 GSLSKLSGKGLQRMFFLNDVLLYTSRGLTASNQPKVHGQLPLYGMTIESEDEWGVPH 60  
 Db 764 gsisklskgqlqrmfflndvlltsrgltasnkqkvhgqlplygmtieesedewgvph 823  
 QY 61 CLTLRGQRQSIIVAAASRSRSEMEKWWEDIQMA 91  
 Db 824 cltlrgqrqsiivaaasrsrsemekwwediqma 854

RESULT 3  
 AAY91947  
 ID AAY91947 standard; Protein; 1045 AA.  
 XX  
 AC AAY91947;  
 XX  
 DT 19-JUL-2000 (first entry)  
 XX  
 DE Human cytoskeleton associated protein 2 (CYSKP-2).  
 XX  
 KW Cytoskeleton associated protein; CYSKP-2; cancer; proliferative;  
 KW autoimmunity; inflammatory; vesicle trafficking; neurological;  
 KW cardiovascular; cell motility; reproductive; muscle disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Modified-site 23  
 FT /note= "potential phosphorylation site"  
 FT Modified-site 36  
 FT /note= "potential phosphorylation site"  
 FT Modified-site 41  
 FT /note= "potential phosphorylation site"  
 FT Domain 47..85  
 FT /note= "signature sequence"  
 FT Modified-site 92  
 FT /note= "potential phosphorylation site"  
 FT Domain 94..123  
 FT /note= "signature sequence"  
 FT Domain 144..190  
 FT /note= "signature sequence"  
 FT Modified-site 150  
 FT /note= "potential phosphorylation site"  
 FT Modified-site 152  
 FT /note= "potential N-glycosylation site"  
 FT Domain 196..249  
 FT /note= "signature sequence"  
 FT Modified-site 207  
 FT /note= "potential phosphorylation site"  
 FT Domain 261..279  
 FT /note= "signature sequence"





```
XX 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX
XX 27-AUG-1999; 99JP-0300253.
XX
XX 11-JAN-2000; 2000JP-0118776.
XX
XX 02-MAY-2000; 2000JP-0183767.
XX
XX 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
XX full-length cDNAs defined in the specification, and for the detection
XX and/or diagnosis of the abnormality of the proteins encoded by the
XX full-length cDNAs -
XX
XX Claim 8; SEQ ID 12972; 2537pp + CD ROM; English.
XX
XX The present invention describes primer sets for synthesising 5602
XX full-length cDNAs defined in the specification. Where a primer set
XX comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX to the complementary strand of a polynucleotide which comprises one of
XX the 5602 nucleotide sequences defined in the specification, where the
XX oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX of an oligonucleotide comprising a sequence complementary to the
XX complementary strand of a polynucleotide which comprises a 5'-end
XX sequence and an oligonucleotide comprising a sequence complementary to a
XX polynucleotide which comprises a 3'-end sequence, where the
XX oligonucleotide comprises at least 15 nucleotides and the combination
XX of the 5'-end sequence/3'-end sequence is selected from those defined in
XX the specification. The primer sets can be used in antisense therapy and
XX in gene therapy. The primers are useful for synthesising polynucleotides,
XX particularly full-length cDNAs. The primers are also useful for the
XX detection and/or diagnosis of the abnormality of the proteins encoded by
XX the full-length cDNAs. The primers allow obtaining of the full-length
XX cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX represent oligonucleotides, all of which are used in the exemplification
XX of the present invention.
XX
XX Sequence 432 AA;
XX
XX Query Match 32.6%; Score 152.5; DB 22; Length 432;
XX Best Local Similarity 40.7%; Pred. No. 2,9e-12;
XX Matches 37; Conservative 16; Mismatches 35; Indels 3; Gaps 2;
XX
XX 1 GSKSLGKGLQQRMFLENDVLYTSRGLTASNQFKVHGQQLPLYGMTIESEDEWGVPH 60
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 96 gilmklsrkvmqpmrffldallytpp--vqsgmyklnmlslagmkvrkptqe-ayqn 152
XX
XX 61 CLTLRGQRQSIIVAASSRSEMEKWEVDIQNA 91
XX | : : : : : | : : : : : | : : : : : |
XX 153 elkiesversflissassaterdewlealsra 183
XX
XX RESULT 5
XX AAY51248
XX ID AAY51248 standard; Protein; 766 AA.
XX
XX AC AAY51248;
XX
XX 14-APR-2000 (first entry)
XX
XX DE Rat actin-binding protein frabin.
XX
XX Actin-binding protein; frabin; rat; developmental stage; animal growth.
```

```
XX Rattus sp.
XX
XX JP11346775-A.
XX
XX 21-DEC-1999.
XX
XX 11-JUN-1998; 98JP-0164016.
XX
XX 11-JUN-1998; 98JP-0164016.
XX
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
XX (OBAI/) OBAISHI H.
XX
XX WPI; 2000-109691/10.
XX N-PSDB; AA244678.
XX
XX An actin-binding protein frabin - plays an important role in the
XX developmental stages of animal growth
XX
XX Claim 1; Page 6-8; 12pp; Japanese.
XX
XX This invention describes a novel actin-binding protein, frabin (I). The
XX new actin-binding protein frabin plays an important role in the
XX developmental stages of animal growth. The protein can be used in gene
XX engineering. This sequence represents the rat frabin protein described
XX in the method of the invention.
XX
XX Sequence 766 AA;
XX
XX Query Match 26.1%; Score 122; DB 21; Length 766;
XX Best Local Similarity 31.5%; Pred. No. 1e-07;
XX Matches 29; Conservative 24; Mismatches 35; Indels 4; Gaps 3;
XX
XX 1 GSKSLGK--GLQQRMFLENDVLY--TSRGLTASNQFKVHGQQLPLYGMTIESEDEW 57
XX | : | : | : | | | | | | | | | | | | | | | | | | | | | | | |
XX 427 gqilkaarntsagerylflfnmnllycvprfslvsgskftvtrvgidgmknivethne-e 485
XX
XX 58 VPCHLTLRGQRQSIIVAASSRSEMEKWEVDIQ 89
XX | | : : : : : | | : : : : : | | : : : : : |
XX 486 yphftvgsgkrtlelqasseqdkewikalq 517
XX
XX RESULT 6
XX AAM39396
XX ID AAM39396 standard; Protein; 183 AA.
XX
XX AC AAM39396;
XX
XX 22-OCT-2001 (first entry)
XX
XX DE Human polypeptide SEQ ID NO 2541.
XX
XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
XX peripheral nervous system; neuropathy; central nervous system; CNS;
XX Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
XX anyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
XX chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
XX leukaemia.
XX
XX Homo sapiens.
XX
XX WO200153312-A1.
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX
XX 25-APR-2000; 2000US-0552317.
XX
XX 09-JUL-2000; 2000US-0598042.
XX
XX 19-JUL-2000; 2000US-0620312.
```









XX (HELI-) HELIX RES INST.

XX Ota T, Isoqai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI: 2001-318749/34.

XX

PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

XX

XX Claim 8; SEQ ID 14968; 2537pp + CD ROM; English.

PS

XX The present invention describes primer sets for synthesising 5602

CC full-length cDNAs defined in the specification. Where a primer set

CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary

CC to the complementary strand of a polynucleotide which comprises one of

CC the 5602 nucleotide sequences defined in the specification, where the

CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination

CC of an oligonucleotide comprising a sequence complementary to the

CC complementary strand of a polynucleotide which comprises a 5'-end

CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the

CC oligonucleotide comprises at least 15 nucleotides and the combination of

CC the 5'-end sequence/3'-end sequence is selected from those defined in

CC the specification. The primer sets can be used in antisense therapy and

CC in gene therapy. The primers are useful for synthesising polynucleotides

CC particularly full-length cDNAs. The primers are also useful for the

CC detection and/or diagnosis of the abnormality of the proteins encoded by

CC the full-length cDNAs. The primers allow obtaining of the full-length

CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and

CC AAH13633 to AAH18742 represent human cDNA sequences; AAB952446 to

CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632

CC represent oligonucleotides, all of which are used in the exemplification

CC of the present invention.

XX

XX Sequence 762 AA;

SQ

Query Match 18.3%; Score 85.5; DB 22; Length 762;

Best Local Similarity 23.4%; Pred. No. 0.011;

Matches 25; Conservative 23; Mismatches 40; Indels 19; Gaps 3

QY 1 GSLSKSLGKGLQRMFFLENDVLLY-----TSRGLTASNQPKVHGQLPLYGNTIEE 51

Db | | | | : | : | | | : | : | : | : | : | : | : | : | : |

33 gvlkissnqivgfifdlvlyckrkhrrlknksastdghrylfrgrintevmeven 92

QY 52 SEDWGVPHC---LTLRG-----QRQSIIVAASRSSEMEKQVEDI 88

Db | | | | : | : | : | : | : | : | : | : | : |

93 vddgtadfsghivngvkihtaknkvfcmaktpeekhewfai 139

RESULT 14

AAW26726

ID AAW26726 standard; Protein; 394 AA.

XX

AC AAW26726;

XX

DT 27-APR-1998 (first entry)

XX

XX Human integrin regulatory protein IRP-2.

DE

XX IRP-2; integrin regulatory protein 2; human; cytoplasmic modulator;

KW signal transduction; inflammation; therapy.

XX

OS Homo sapiens.

XX

XX Location/Qualifiers

FT 9..60

FT Domain

FT /note= "kinesin/myosin domain"

FT Domain

FT 71..245

FT

FT	Domain	/note= "SEC7 domain"
FT		260..394
FT		/note= "pleckstrin homology domain"
XX	WO7939124-Al.	
XX	23-OCT-1997.	
XX	15-APR-1997;	97WO-US06272.
XX	15-APR-1996;	96US-0632247.
XX	(ICOS-) ICOS CORP.	
XX	Lipsky BP, Staunton DE;	
XX	WPI: 1997-526461/48.	
XX	N-PSDB; AAT99408.	
XX	DNA encoding human integrin regulatory proteins 1 and 2 - useful for identifying cytoplasmic modulators of integrin regulation or signalling.	
XX	Claim 2; Page 59-60; 82pp; English.	
XX	This protein comprises human integrin regulatory protein 2 (IRP-2), a protein which regulates beta-2 and beta-7 integrins and which is contemplated to participate in integrin signalling and/or recycling pathways. Its amino acid sequence was deduced from a human spleen cDNA clone (see AAT99408). Also claimed are: a DNA expression construct; a transformed or transfected host cell; purified IRP-1 (see AAM33417) and IRP-2 polypeptides; antibodies, preferably monoclonal, which specifically bind IRP-1 or IRP-2; an anti-idiotypic antibody; hybridoma cell lines; a method for identifying a compound that modulates binding between IRP-1 or IRP-2 and a beta integrin subunit, ADP ribosylation factor or phosphatidylinositol; and a method for isolating a polynucleotide encoding a protein that binds to IRP-1 or -2 comprising the yeast two-hybrid system. Modulators e.g. IRP binding proteins, of IRP interactions can be used in monitoring and treating inflammatory processes involving leukocytes.	
XX	Sequence	394 AA;
XX	Query Match	18.1%; Score 84.5; DB 18; Length 394;
XX	Best Local Similarity	25.7%; Pred. No. 0.0059;
XX	Matches	29; Conservative
XX	Mismatches	22; Indels
XX	Gaps	29; Gaps
QY	1	GSLSKLSGG--KGIQRMFFLNFMDVLLYTSRGLTASNOFKVHGOLPLYGMTIESEDEMGV 58
Db	264	gwlkiggrvktwrwfltdnclyy----feftdkegrgiplenisvqkvddp-kk 31
QY	59	PHCLTL-----RGQR-----QSIIVASRSRSEMKKWEVDIQ 89
Db	319	pfcllelynpscrgqkackictdgdgrvvvgkhesyrisatsaerdqgiesir 371
RESULT	15	
AAAY26928		
ID	RAY26928	standard; Protein; 394 AA.
XX	AAAY26928;	
XX	21-DEC-1999	(first entry)
XX	Human integrin regulatory protein 2.	
XX	Human; integrin regulatory protein 1; IRP-1; modulation; IRP-2;	
XX	beta-integrin; vascular cell adhesion molecule 1; VCAM-1; ARF; PI;	
XX	ADP ribosylation factor; phosphatidylinositol; inflammation; leukoc	
XX	Homo sapiens.	

PN US958705-A.

XX

PD 28-SEP-1999.

XX

XX 15-APR-1997; 97US-0839581.

XX

XX 15-APR-1996; 96US-0632247.

XX

XX (ICOS-) ICOS CORP.

XX

XX Liskay BP, Staunton DE;

XX

XX WPI; 1999-570764/48.

XX

XX N-ESDB; RAZ24909.

XX

PT Identifying modulators of integrin regulatory protein interactions

XX

PS Disclosure; Column 35-38; 17pp; English.

XX

This sequence represents a human integrin regulatory protein 2 (IRP-2). The protein is used in a method of identifying compounds that modulate binding of IRP-1 (AAZ24908) or IRP-2 to a beta-integrin subunit. IRP-1 and IRP-2 proteins are characterized by the ability to abrogate alpha-4-beta-7 binding to vascular cell adhesion molecule 1 (VCAM-1) in JY cells. The proteins can also be used to identify compounds that modulate binding between IRP-1 or IRP-2 and ADP ribosylation factor (ARF) or phosphatidylinositol (PI). Modulators of ARF, PI, beta1, beta2, beta3 or beta7/IRP interaction may be used in vitro or in vivo to affect inflammatory processes involving leukocytes.

XX Sequence 394 AA;

SQ

Query Match 18.1%; Score 84.5; DB 20; Length 394;

Best Local Similarity 25.7%; Pred. No. 0.0059;

Matches 29; Conservative 22; Mismatches 33; Indels 29; Gaps 5;

Qy 1 GSLSKLGG--KGLQRMFFLNDVLLYTSRGLTASNQPKVHGQLPLYGMTIESEDEWGV 58

| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 264 gwlklgrvktwkrwfiltdnclyy---feftdkeprgiiplenlsvgkvddp-kk 318

Qy 59 PHCLTL-----RGQR-----QSIIVAASSRSEMEKWWEDIQ 89

| | | | | : | | | | | : | | | | | : | | | | | : | | | | |

Db 319 pfclelynpsergqkacktdgdgrvvvghkhesyrisatsaeerdqwiesir 371

Search completed: December 6, 2001, 08:51:44  
Job time: 336 sec